
Canterbury
Christ Church
University

Synchronous diversification of Sulawesi's iconic artiodactyls driven by recent geological events

Authors

Laurent A. F. Frantz[1,2,a,*], Anna Rudzinski[3,*], Abang Mansyursyah Surya Nugraha[4,c,*], , Allowen Evin[5,6*], James Burton[7,8*], Ardern Hulme-Beaman[2,6], Anna Linderholm[2,9], Ross Barnett[2,10], Rodrigo Vega[11] Evan K. Irving-Pease[2], James Haile[2,10], Richard Allen[2], Kristin Leus[12,13], Jill Shephard[14,15], Mia Hillyer[14,16], Sarah Gillemot[14], Jeroen van den Hurk[14], Sharron Ogle[17], Cristina Atofanei[11], Mark G. Thomas[3], Friederike Johansson[18], Abdul Haris Mustari[19], John Williams[20], Kusdiantoro Mohamad[21], Chandramaya Siska Damayanti[21], Ita Djuwita Wiryadi[21†], Dagmar Obbles[22], Stephano Mona[23,24], Hally Day[25], Muhammad Yasin[25], Stefan Meker[26] , Jimmy A. McGuire[27], Ben J. Evans[28], Thomas von Rintelen[29], Simon Y. W. Ho[30], Jeremy B. Searle[31], Andrew C. Kitchener[32,33], Alastair A. Macdonald[7b], Darren J. Shaw[7b], Robert Hall[4,b], Peter Galbusera[14,b] and Greger Larson[2,a,b]

[1] School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK

[2] The Palaeogenomics & Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, University of Oxford, Oxford OX1 3QY, UK

[3] Research Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK

[4] SE Asia Research Group, Department of Earth Sciences, Royal Holloway University of London, Egham, Surrey, TW20 0EX, UK

[5] Institut des Sciences de l'Evolution, Université de Montpellier, CNRS, IRD, EPHE, Place Eugène Bataillon, 34095 Montpellier Cedex 05, France

[6] Department of Archaeology, Classics and Egyptology, University of Liverpool, 12-14 Abercromby Square, Liverpool, L69 7WZ, UK

[7] Royal (Dick) School of Veterinary Studies & The Roslin Institute, University of Edinburgh, Easter Bush Campus, Roslin, Edinburgh EH25 9RG, UK

[8] IUCN SSC Asian Wild Cattle Specialist Group and Chester Zoo, Cedar House, Caughall Road, Upton by Chester, Chester CH2 1LH, UK

[9] Department of Anthropology, Texas A&M University, College Station, TX 77843-4352, USA.

[10] Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen K, Denmark

[11] Ecology Research Group, Section of Life Sciences, School of Human and Life Sciences, Canterbury Christ Church University, North Holmes Road, Canterbury, CT1 1QU, Kent, UK

[12] Copenhagen Zoo, IUCN SSC Conservation Breeding Specialist Group - Europe, Roskildevej 38, Postboks 7, DK-2000 Frederiksberg, Denmark

[13] European Association of Zoos and Aquaria, PO Box 20164,1000 HD Amsterdam, The Netherlands

[14] Centre for Research and Conservation (CRC), Royal Zoological Society of Antwerp, Koningin Astridplein 20-26, 2018 Antwerp, Belgium.

[15] Environment and Conservation Sciences, School of Veterinary and Life Sciences, Murdoch University, Perth, WA 6150, Australia.

[16] Molecular Systematics Unit / Terrestrial Zoology, Western Australian Museum, Welshpool, WA, Australia

[17] Edinburgh Medical School: BMTO, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, UK.

[18] Gothenburg Natural History Museum, Box 7283, S 402 35 Gothenburg, Sweden

[19] Department of Forest Resources Conservation and Ecotourism, Faculty of Forestry, Bogor Agricultural University, PO Box 168, Bogor 16001, Indonesia

[20] Davies Research Centre, School of Animal and Veterinary Sciences, Faculty of Sciences, University of Adelaide, Roseworthy, SA 5371, Australia

[21] Faculty of Veterinary Medicine, Bogor Agricultural University, Jalan Agatis, IPB Campus Darmaga Bogor 16680, Indonesia

[22] Laboratory of Aquatic Ecology, Evolution and Conservation, KU Leuven, Ch. Deberiotstraat 32, 3000 Leuven, Belgium

[23] Institut de Systématique, Évolution, Biodiversité, ISYEB - UMR 7205 - CNRS, MNHN, UPMC, EPHE, Ecole Pratique des Hautes Etudes, 16 rue Buffon, CP39, 75005, Paris, France

[24] EPHE, PSL Research University, Paris, France

[25] No affiliation

[26] Department of Zoology, State Museum of Natural History Stuttgart, Rosenstein 1, 70191 Stuttgart, Germany

[27] Museum of Vertebrate Zoology and Department of Integrative Biology, University of California, Berkeley, CA 94720, USA

[28] Department of Biology, McMaster University, Hamilton, Ontario, Ontario, Canada

[29] Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany

[30] School of Life and Environmental Sciences, University of Sydney, Sydney, NSW 2006, Australia

[31] Department of Ecology and Evolutionary Biology, Corson Hall, Cornell University, Ithaca, NY 14853, USA

[32] Department of Natural Sciences, Chambers Street, National Museums Scotland, Edinburgh EH1 1JF, UK.

[33] Institute of Geography, School of Geosciences, Drummond Street, University of Edinburgh, Edinburgh EH8 9XP, UK.

[†] deceased
[*]: contributed equally
[b]: co-supervised the study
[a] corresponding authors: laurent.frantz@qmul.ac.uk and greger.larson@arch.ox.ac.uk
[c] Present address: Pertamina University, Jl. Teuku Nyak Arief, Kawasan Simprug, Kebayoran Lama, Jakarta Selatan 12220, Indonesia

Keywords: biogeography, evolution, geology, Wallacea.

Abstract

The high degree of endemism on Sulawesi has previously been suggested to have vicariant origins, dating back 40 Myr ago. Recent studies, however, suggest that much of Sulawesi's fauna assembled over the last 15 Myr. Here, we test the hypothesis that more recent uplift of previously submerged portions of land on Sulawesi promoted diversification, and that much of its faunal assemblage is much younger than the island itself. To do so, we combined palaeogeographical reconstructions with genetic and morphometric data sets derived from Sulawesi's three largest mammals: the Babirusa, Anoa, and Sulawesi warty pig. Our results indicate that although these species most likely colonized the area that is now Sulawesi at different times (14 Myr ago to 2-3 Myr ago), they experienced an almost synchronous expansion from the central part of the island. Geological reconstructions indicate that this area was above sea level for most of the last 4 Myr, unlike most parts of the island. We conclude that emergence of land on Sulawesi (~1–2 Myr) may have allowed species to expand synchronously. Altogether, our results indicate that the establishment of the highly endemic faunal assemblage on Sulawesi was driven by geological events over the last few million years.

Introduction

Alfred Russel Wallace was the first to document the 'anomalous' biogeographic region in Island Southeast Asia now known as Wallacea [1,2]. This biodiversity hotspot [3] is bounded by Wallace's Line in the west and Lydekker's Line in the east [4]. It consists of numerous islands in the Indonesian archipelago, all of which boast a high degree of endemism. For example, on Sulawesi, the largest island in the region, at least 61 of the 63 non-volant mammalian species are endemic [5] and this figure is likely to be an underestimate.

The geological origins of Wallacea are as complex as its biogeography. Until recently, Sulawesi had been regarded as the product of multiple collisions of continental fragments from the Late Cretaceous [6–9]. This assumption has been challenged and a recent reinterpretation suggests instead that the island began

to form as the result of continental collisions during the Cretaceous, which were then followed by Eocene rifting of the Makassar Strait. This process led to the isolation of small land areas in western Sulawesi from Sundaland. In the Early Miocene (~23 Myr ago), a collision between the Sula Spur (a promontory of the Australian continent) and north Sulawesi led to uplift and emergence of land [10–12]. Later tectonic movements led to the present-day configuration of islands between Borneo and Australia [13,14].

Previous geological interpretation involving the assembly of multiple terranes by collision was used to suggest that Sulawesi's peculiar species richness resulted from vicariance and amalgamation over long geological time periods [10,15,16]. However, recent molecular-clock analyses suggest that a dispersal, starting in the middle Miocene (~15 Myr ago) from both Sunda and Sahul is a more plausible explanation [17–19]. These conclusions suggest a limited potential for animal dispersal to Sulawesi prior to ~15 Myr ago. Rapid tectonic changes, coupled with the dramatic sea-level fluctuations over the past 5 Myr [20], might also have affected land availability and influenced patterns of species dispersal to Sulawesi, intra-island species expansion and speciation.

The hypothesis of a recent increase in land area [19] can be tested by comparing the population histories of multiple species on the island. Analyses of genetic and morphometric variability can be used to infer the timing and trajectories of dispersal, and the geographical and temporal origins of expansion. For example, if land area had increased, from a single smaller island, extant species now living on Sulawesi, would all have expanded from the same area. In addition, under this assumption, within the same geographical region their respective diversifications would be expected to have been roughly simultaneous.

Here, we focus on three large mammals endemic to Sulawesi: the Babirusa (*Babyrousa* spp.), the Sulawesi warty pig (SWP, *Sus celebensis*) and the Anoa, a dwarf buffalo, (*Bubalus* spp.). The Babirusa (*Babyrousa spp.*) is a suid characterized by wrinkled skin and two extraordinary curved upper canine tusks

displayed by males [21–23]. It represents a "ghost lineage" since there are no closely related extant species outside Sulawesi (*e.g.* African suids are more closely related to all other Asian suids than to Babirusa) and the Babirusa is unknown in the fossil record outside Sulawesi [24]. Three extant species of Babirusa (distributed primarily in the interior of Sulawesi and on surrounding islands [21–23] have been described: *Babyrousa babyrussa* (Buru and Sulu islands), *Babyrousa celebensis* (mainland Sulawesi) and *Babyrousa togeanensis* (Togian Island) [25].

The Anoa is an endemic "miniature buffalo" related to indigenous bovids in the Philippines and East Asia [26,27]. It stands approximately one metre tall, weighs 150–200 kg, and mostly inhabits pristine rainforest [28]. Although the subgenus *Anoa* has been divided into two species, the lowland Anoa (*Bubalus depressicornis*) and the highland Anoa (*Bubalus quarlesi*) [29], this classification is still contentious [27]. In contrast with Anoa and Babirusa, the Sulawesi warty pig (SWP; *Sus celebensis*) occupies a wide range of habitats, ranging from swamps to rainforests. This species is closely related to the Eurasian wild pig(*Sus scrofa*), from which it diverged during the early Pleistocene (~2 Myr ago)[24,30]. The SWP has been found on numerous islands throughout Island Southeast Asia (ISEA), probably as the result of human-mediated dispersal [31]. As its name implies, male SWPs develop facial warts. These cultural icons (*e.g.* SWP/Babirusa and Anoa are represented in the oldest prehistoric cave paintings [32,33]) have undergone recent and significant population reduction and range contraction due to overhunting and conversion of natural habitat for agricultural use.

Here, we establish when Sulawesi gained its modern shape and size, including connectivity between its constituent peninsulae, and assessed the impact of island formation on the evolution of Sulawesi's biodiversity. To do so, we used new reconstructions of the island's palaeogeography that allowed us to interpret the distribution of land and sea over the last 8 Myr at 1 Myr intervals. To determine the timings of diversification of the three largest endemic mammals on the island, we generated and analysed genetic and/or morphometric data

from a total of 1,289 samples of the SWP, Anoa, and Babirusa obtained from museums, zoos and wild populations (456, 520 and 313 samples respectively; Table S1). More specifically, we measured a total of 356 teeth from 227 specimens (357 Babirusa and 191 SWP) using a geometric morphometric approach. In addition, we sequenced mitochondrial loci (cytb and/or control region) from 142 Anoas, 213 Babirusa and 230 SWP. Lastly, we typed typed 13 microsatellite loci from 163 Anoa, 14 loci from 238 SWP, and 13 from 182 Babirusa (see Electronic Supplementary for more information). Although these taxa have been divided into multiple species (see taxonomic notes in the Electronic Supplementary Material), for the purpose of this study we treated SWP, Anoa and Babirusa as single taxonomic units.

Results and Discussion

Contemporaneous divergence

We generated mitochondrial DNA (mtDNA) sequences and/or microsatellite data from 230 SWPs, 155 Anoas and 213 Babirusas sampled across Sulawesi and the neighbouring islands (Electronic Supplementary Material Figure S1; Table S1). Using a molecular-clock analysis, we inferred the time to the most recent common ancestor (TMRCA) of each species. The estimates from this method represent coalescence times, which provide a reflection of the crown age of each taxon. The closer relationship between Babirusa and SWP (~13 Myr ago) [34], compared with the divergence of either species from the Anoa (~58 Myr ago)[35] allowed us to align sequences from Babirusa and SWP alongside one another and jointly infer their relative TMRCAs. Separate analyses were performed for the Anoa. The inferred TMRCA of SWP was 2.19 Myr (95% credibility interval [CI] 1.19–3.41 Myr; Electronic Supplementary Material Figure S2) and of Babirusa was 2.49 Myr (95% CI 1.33–3.61 Myr) (Figure 1; Electronic Supplementary Material Figure S2). The inferred TMRCA of Anoa was younger (1.06 Myr; Figure 1; Electronic Supplementary Material Figure S3), though its 95% CI (0.81–1.96 Myr) overlapped substantially with the TMRCAs of the other two species.

The relatively recent divergence between Babirusa and SWP also allowed us to compare their TMRCAs using identical microsatellite loci. To do so, we computed the average square distance (ASD)[36,37] between every pair of individuals within each species at the same 13 microsatellite loci. Although such an analysis might be affected by population structure (see below), we found that the distributions of ASD values were not significantly different between these two species (Wilcoxon signed-rank test, p=0.492). This is consistent with the mitochondrial evidence for the nearly identical TMRCAs in the two species.

Recent molecular analyses have indicated that Babirusa may have colonized Wallacea as early as 13 Myr ago, whereas SWP and Anoa appear to have only colonized Sulawesi within the last 2–4 Myr [17,30,32,34]. An early dispersal of Babirusa to Sulawesi (late Palaeogene) has also been suggested on the basis of palaeontological evidence [19]. In addition, our data corroborate previous studies in indicating that both SWP and Babirusa are monophyletic with respect to their most closely related taxa on neighbouring islands (*e.g.* Borneo), which is consistent with only one colonization of Sulawesi (Electronic Supplementary Material; Figure S4-6)[30].

We then examined whether patterns of morphological diversity in these taxa are consistent with the molecular date estimates. To do so, we obtained measurements of 356 second and third lower molar (M2 and M3) from 95 Babirusas and 132 SWPs. SWP and Babirusa do not overlap morphologically (Figure 2a) and we were thus able to assign each specimen to its correct species with a success rates of 94.3% (CI: 92.7%–95.5%, distribution of leave-one-out cross validation of a discriminant analysis based on a balanced sample design) [38] and 94.7% (CI: 93.8%–96.7%) based on their M2 and M3, respectively. Our results also indicate that Babirusa did not accumulate more tooth shape variation within Sulawesi (Fligner-Killeen test X^2=1.04, p=0.3 for M2, X^2=3.45, p=0.06 for M3). The data instead suggests that SWP has greater variance in the size of its M3 (X^2=4.52, p=0.03, but not in the size of the M2, X^2=3.44, p=0.06), and that the population from West Central Sulawesi has an overall smaller tooth

size than the two populations from North West and North East Sulawesi (Figure 2b, Table S2). While these results may result from different selective constraints, they indicate that Babirusa did not accumulate greater morphological variation in tooth shape than did the SWP, despite arriving on Sulawesi up to 10 Myr earlier.

Altogether our analyses suggest that although the three species are believed to have colonized the island at different times, their similar degrees of morphological diversity and their nearly synchronous TMRCAs raise the possibility that they (and possibly other species) responded to a common mechanism that triggered their contemporaneous diversification.

Past land availability correlates with the expansion origins

Increasing land area may have promoted a simultaneous diversification and range expansion in Babirusas, SWPs, and Anoas. To test this hypothesis, we used a new reconstruction that depicts land area in the Sulawesi region through time using information from the geological record. The reconstructions in 1 Myr increments (Figure 3a; Figure S7; [39]) support a scenario in which most of Sulawesi was submerged until the late Pliocene to early Pleistocene (2–3 Myr ago). Large-scale uplifts over the last 2–3 Myr would have rapidly and significantly increased land area, making it possible for non-volant species to expand their ranges.

To further assess whether these Plio-Pleistocene uplifts were responsible for a synchronous expansion, we inferred the most likely geographical origin of expansion using microsatellite data under a model of spatial loss-of-diversity with distance from expansion origin (Electronic Supplementary Material). These estimates were obtained independently of, and uninformed by, either the geological reconstructions or modern phylogeographical boundaries inferred from other species. We deduced that the most likely origin for both SWP and Babirusa was in the East Central region of Sulawesi (Figure 3c and 3d), and the most likely origin of Anoa was in the West Central region (Figure 3b).

The origins of the population expansions of both SWP and Babirusa occurred in an area of Sulawesi that only emerged during the late Pliocene to early Pleistocene (Figure 3a; Electronic Supplementary Material Figure S7). On the other hand, the Anoa most likely origin of diversification lies in a region that was submerged until the Pleistocene, consistent with paleontological evidence [32] and with the slightly more recent TMRCA inferred for this species (Figure 1). Thus, for all three species, the inferred geographical origins of their range expansions match the land availability derived from our geological reconstruction of Sulawesi.

Geological history of past land isolation correlates with zones of endemism

Previous studies have identified endemic zones that are common to macaques, toads [18,40], tarsiers [41–44] and lizards [45]. We tested whether the same areas of endemism are linked to the population structure in our three species by generating a phylogenetic tree for each species using mtDNA and defined 5–6 haplogroups per species based on well-supported clades (Figure 4a-c; *Electronic Supplementary Material* Figure S4-6). We found that haplogroup proportions were significantly different between previously defined areas of endemism in all three species (Pearson's chi-squared test*;* $p<0.001$), suggesting population substructure.

We also used STRUCTURE [46] to infer population structure from microsatellite data. The optimum numbers of populations (K) were 5, 6 and 5 for Anoa, Babirusa and SWP, respectively (Electronic Supplementary Material Figure S8; Figure 4d-f). Plotting the proportion of membership of each sample onto a map revealed a strong correspondence with the previously described zones of endemism (Figure 4d-f). Using an analysis of molecular variance (AMOVA), we found that these areas of endemism explained approximately 17%, 27%, and 5% of the variance in allele frequencies in Anoa, Babirusa and SWP, respectively (Table S5). Populations of Babirusa and SWP in these zones of endemism were also strongly morphologically differentiated (Figure 2).

Altogether, these data and analyses indicate that, despite some differences, the zones of endemism identified in tarsiers, macaques, toads and lizards [18,40–45,47] are largely consistent with the population structure and morphological differentiation in the three species studied here. This is particularly striking for the north arm of Sulawesi (NW, NC, and NE in Figure 4), where we identify two highly differentiated populations (reflected in both mtDNA and nuclear data sets) in all three taxa. This pattern could result from either adaptation to local environments or from isolation due to the particular geological history associated with the northern arm. Geological reconstructions (Figure 3a) indicate that although land was present in this region during the past 4 Myr, it was often isolated from the rest of Sulawesi until the mid-Pleistocene. Thus, the combined geological and biological evidence presented here indicate that the high degree of divergence observed in the northern-arm populations in a multitude of species (*e.g.* three ungulates, macaques, and tarsiers) might have been shaped by isolation from the rest of the island until the last 1My (Figure 3a)
.

Recent and contemporary land isolation also affected morphological evolution including dwarfism

Similar isolation is likely to have influenced the populations inhabiting the smaller islands adjacent to Sulawesi, including the Banggai archipelago, Buru, the Togian and Sula Islands. Interestingly, our geometric morphometric analyses demonstrated that these island populations of SWP and Babirusa are the most morphologically divergent (Figure 2a). For example, the insular populations from the Togian Islands (Babirusa) and the Banggai archipelago (SWP) were found to have much smaller tooth sizes than their counterparts on the mainland (Figure 2b).

The significant morphometric divergences between populations on various islands are consistent with the genetic differentiation between Babirusa/SWP on Togian, Sula, and Buru (Figure 4; *Electronic Supplementary Material* Figure S9;

Electronic Supplementary Material Figure S10) and between island populations of SWP on Banggai archipelago, Buton, and Buru (Figure 4; *Electronic Supplementary Material* Figure S9; *Electronic Supplementary Material* Figure S10).

Together, these results show that while suture zones between tectonic fragments are consistent with genetic and morphometric differentiation within Sulawesi, isolation on remote islands is likely to have had a much greater effect on morphological distinctiveness. Rapid evolution, on islands, has been described in many species (e.g [48]) including in pigs [49] where island populations are known to have smaller tooth sizes than their mainland counterparts [50,51].

Demographic history

Isolation of subpopulations across Sulawesi might also be linked to recent anthropogenic disturbances, especially for Anoa and Babirusa, that occupy pristine forest or swamps [21,28]. In order to assess the impact of recent anthropogenic changes on the three species, we inferred their demographic history using approximate Bayesian computation (ABC). We fitted various demographic models to the genetic data (combining both mtDNA and microsatellite data; *Electronic Supplementary Material;* Figure S11). The best-supported demographic model involved a long-term expansion followed by a recent bottleneck in all three species (Table S3), corroborating the results of recent analyses of the SWP genome [30].

While our ABC analysis had insufficient power to retrieve the time of expansion (Table S4), it provided relatively narrow estimates of the current effective population sizes (Figure 5; Table S4). We inferred a larger effective population size in SWP (83,021; 95% CI 46,287–161,457) than in Babirusa (30,895; 95% CI 17,522–54,954) or Anoa (27,504; 95% CI 13,680–54,056). *Sus celebensis* occupies a wide range of habitats, including agricultural areas [52]. Thus, this species is likely to be less affected by continuing deforestation than Babirusa or Anoa, which are typically restricted to less disturbed forest and swamps [21,26].

Phylogenetic analyses of microsatellite data indicate more geographical structuring in Babirusa and Anoa than in SWP (*Electronic Supplementary Material* Figure S12; Table S5). Altogether, these results are consistent with species-specific responses to habitat loss.

Conclusions

Our results indicate that, while the different geological components of Sulawesi were assembled at about 23 Myr ago, the island only acquired its distinctive modern form in the last few million years. By 3 Myr ago there was a large single island at its modern centre, but the complete connection between the arms was established more recently. The increasing land area associated with Plio-Pleistocene tectonic activity is likely to have provided the opportunity for a synchronous expansion in the three endemic mammal species in this study, as well as numerous other species. Interestingly, both our Pleistocene geological reconstruction and our proposed origins of expansion in the centre of the island closely resemble maps inferred from a study of tarsier species distribution on Sulawesi [42].

Furthermore, the recent emergence of connections between Sulawesi's arms coincides with a faunal turnover on the island and the extinction of multiple species. The geological reconstruction, and in particular the recent elimination of the marine barrier at the Tempe depression separating the Southwest and Central regions, fits well with suggested replacement in tarsier species that occurred in the last ~1 My [41]. The dispersal of our three species from the central region of Sulawesi may therefore have played a role in other local extinctions, such as the extinct suid known from Southwest Sulawesi, Celebochoerus.

Sulawesi's development by emergence and coalescence of islands had a significant impact on the population structure and intraspecific morphological differentiation of Sulawesi's three largest mammals and many other endemic taxa. Thus, while most of Sulawesi's extant fauna arrived relatively recently, the

more ancient geological history of the island (collision of multiple fragments) might have also affected patterns of endemism. Many aspects of Sulawesi's interconnected natural and geological histories remain unresolved. Integrative approaches that combine biological and geological data sets are therefore essential for reconstructing a comprehensive evolutionary history of Wallace's most anomalous island.

References

1. Wallace AR. 1863 On the Physical Geography of the Malay Archipelago. *Journal of the Royal Geographical Society of London* **33**, 217.

2. Wallace AR. 2012 ISLAND LIFE. In *Island Life* (ed AR Wallace), pp. xix–xx. Cambridge: Cambridge University Press.

3. Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GA, Kent J. 2000 Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858.

4. Lohman DJ, de Bruyn M, Page T, von Rintelen K, Hall R, Ng PKL, Shih H-T, Carvalho GR, von Rintelen T. 2011 Biogeography of the Indo-Australian Archipelago. *Annu. Rev. Ecol. Evol. Syst.* **42**, 205–226.

5. Musser GG. 1987 The mammals of Sulawesi. *Biogeographical evolution of the Malay Archipelago* , 73–93.

6. Smith RB, Silver EA. 1991 Geology of a Miocene collision complex, Buton, eastern Indonesia. *Geol. Soc. Am. Bull.* **103**, 660–678.

7. Hall R. 1996 Reconstructing Cenozoic SE Asia. In *Tectonic Evolution of SE Asia* (eds R Hall, DJ Blundell), pp. 153–184.

8. Hamilton W. 1979 Tectonics of the Indonesian Region: Geological Survey Professional Paper 1078, US. *Government Printing Office*

9. Davidson JW. 1991 The geology and prospectivity of Buton Island, Southeast Sulawesi, Indonesia. *Indonesian Petroleum Association, Proceedings 20th annual convention, Jakarta 1991* **I**, 209–234.

10. Moss SJ, Wilson MEJ. 1998 Biogeographic implications from the Tertiary palaeogeographic evolution of Sulawesi and Borneo. *In Hall, R. & Holloway, J. D. (eds.) Biogeography and Geological Evolution of SE Asia. Backhuys Publishers, Leiden, The Netherlands* , 133–163.

11. Hall R. 1998 The plate tectonics of Cenozoic SE Asia and the distribution of land and sea. *Biogeography and geological evolution of SE Asia*

12. Hall R. 2013 The palaeogeography of Sundaland and Wallacea since the Late Jurassic. *J. Limnol.* **72**, 1–17.

13. Spakman W, Hall R. 2010 Surface deformation and slab–mantle interaction during Banda arc subduction rollback. *Nat. Geosci.* **3**, 562–566.

14. Hall R. 2011 Australia-SE Asia collision: plate tectonics and crustal flow. In *The SE Asian gateway: history and tectonics of Australia-Asia collision* (eds R Hall, MA Cottam, MEJ Wilson), pp. 75–109.

15. Michaux B. 1996 The origin of southwest Sulawesi and other Indonesian terranes: a biological view. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **122**, 167–183.

16. de Boer AJ, Duffels JP. 1996 Historical biogeography of the cicadas of Wallacea, New Guinea and the West Pacific: a geotectonic explanation. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **124**, 153–177.

17. Stelbrink B, Albrecht C, Hall R, von Rintelen T. 2012 THE BIOGEOGRAPHY OF SULAWESI REVISITED: IS THERE EVIDENCE FOR A VICARIANT ORIGIN OF TAXA ON WALLACE'S 'ANOMALOUS ISLAND'? *Evolution* **66**, 2252–2271.

18. Evans BJ, Supriatna J, Andayani N, Setiadi MI, Cannatella DC, Melnick DJ. 2003 Monkeys and toads define areas of endemism on Sulawesi. *Evolution* **57**, 1436–1443.

19. van den Bergh GD, de Vos J, Sondaar PY. 2001 The Late Quaternary palaeogeography of mammal evolution in the Indonesian Archipelago. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **171**, 385–408.

20. Miller KG *et al.* 2005 The Phanerozoic record of global sea-level change. *Science* **310**, 1293–1298.

21. Macdonald AA, Burton JA, Leus K. 2008 Babyrousa babyrussa. *The IUCN Red List of Threatened Species* (doi:10.2305/IUCN.UK.2008.RLTS.T2461A9441445.en)

22. Macdonald, A., Leus, K., Masaaki, I. & Burton, J. 2016 Babyrousa togeanensis. *IUCN Red List of Threatened Species*. (doi:10.2305/iucn.uk.2016-1.rlts.t136472a44143172.en)

23. Leus, K., Macdonald, A., Burton, J. & Rejeki, I. 2016 Babyrousa celebensis. *IUCN Red List of Threatened Species*. (doi:10.2305/iucn.uk.2016-1.rlts.t136446a44142964.en)

24. Frantz L, Meijaard E, Gongora J, Haile J, Groenen MAM, Larson G. 2016 The Evolution of Suidae. *Annual review of animal biosciences* **4**, 61–85.

25. Meijaard E, Groves C. 2002 Upgrading three subspecies of babirusa (Babyrousa sp.) to full species level. *Asian Wild Pig News* **2**, 33–39.

26. Semiadi G, Mannullang B, Burton JA, Schreiber A, Mustari AH. 2008 Bubalus depressicornis. *The IUCN Red List of Threatened Species* (doi:10.2305/IUCN.UK.2008.RLTS.T3126A9611738.en)

27. Burton JA, Hedges S, Mustari AH. 2005 The taxonomic status, distribution and conservation of the lowland anoa Bubalus depressicornis and mountain anoa Bubalus quarlesi. *Mamm. Rev.* **35**, 25–50.

28. Burton JA, Wheeler P, Mustari A. 2016 Bubalus depressicornis. *The IUCN Red List of Threatened Species* (doi:10.2305/IUCN.UK.2016-2.RLTS.T3126A46364222.en)

29. Groves CP. 1969 Systematics of the anoa (Mammalia, Bovidae). *Beaufortia* **17**, 1–12.

30. Frantz LA *et al.* 2013 Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus. *Genome Biol.* **14**, R107.

31. Groves CP. 1984 Of mice and men and pigs in the Indo-Australian Archipelago. *Canberra Anthropology* **7**, 1–19.

32. Rozzi R. 2017 A new extinct dwarfed buffalo from Sulawesi and the evolution of the subgenus Anoa: An interdisciplinary perspective. *Quat. Sci. Rev.* **157**, 188–205.

33. Aubert M *et al.* 2014 Pleistocene cave art from Sulawesi, Indonesia. *Nature* **514**, 223–227.

34. Gongora J *et al.* 2011 Rethinking the evolution of extant sub-Saharan African suids (Suidae, Artiodactyla). *Zool. Scr.* **40**, 327–335.

35. dos Reis M, Inoue J, Hasegawa M, Asher RJ, Donoghue PCJ, Yang Z. 2012 Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. Biol. Sci.* **279**, 3491–3500.

36. Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW. 1995 An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**, 463–471.

37. Sun JX, Mullikin JC, Patterson N, Reich DE. 2009 Microsatellites are molecular clocks that support accurate inferences about history. *Mol. Biol. Evol.* **26**, 1017–1027.

38. Evin A, Cucchi T, Cardini A, Strand Vidarsdottir U, Larson G, Dobney K. 2013 The long and winding road: identifying pig domestication through molar size and shape. *J. Archaeol. Sci.* **40**, 735–743.

39. Abang Mansyursyah Surya Nugraha and Robert Hall. In press. Late Cenozoic palaeogeography of Sulawesi, Indonesia. *Palaeogeogr. Palaeoclimatol. Palaeoecol.*

40. Evans BJ, Supriatna J, Andayani N, Melnick DJ. 2003 Diversification of Sulawesi macaque monkeys: decoupled evolution of mitochondrial and autosomal DNA. *Evolution* **57**, 1931–1946.

41. Driller C, Merker S, Perwitasari-Farajallah D, Sinaga W, Anggraeni N, Zischler H. 2015 Stop and Go - Waves of Tarsier Dispersal Mirror the Genesis of Sulawesi Island. *PLoS One* **10**, e0141212.

42. Merker S, Driller C, Perwitasari-Farajallah D, Pamungkas J, Zischler H. 2009 Elucidating geological and biological processes underlying the diversification of Sulawesi tarsiers. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 8459–8464.

43. Burton JA, Nietsch A. 2010 Geographical Variation in Duet Songs of Sulawesi Tarsiers: Evidence for New Cryptic Species in South and Southeast Sulawesi. *Int. J. Primatol.* **31**, 1123–1146.

44. Shekelle M, Meier R, Wahyu I, Ting N, Others. 2010 Molecular phylogenetics and chronometrics of Tarsiidae based on 12S mtDNA haplotypes: evidence for Miocene origins of crown tarsiers and numerous species within the Sulawesian clade. *Int. J. Primatol.* **31**, 1083–1106.

45. McGuire JA, Linkem CW, Koo MS, Hutchison DW, Lappin AK, Orange DI, Lemos-Espinal J, Riddle BR, Jaeger JR. 2007 Mitochondrial introgression and incomplete

lineage sorting through space and time: phylogenetics of crotaphytid lizards. *Evolution* **61**, 2879–2897.

46. Pritchard JK, Stephens M, Donnelly P. 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

47. Evans BJ, McGuire JA, Brown RM, Andayani N, Supriatna J. 2008 A coalescent framework for comparing alternative models of population structure with genetic data: evolution of Celebes toads. *Biol. Lett.* **4**, 430–433.

48. Geer A van der. 2010 *Evolution of island mammals : adaptation and extinction of placental mammals on islands*. Wiley-Blackwell.

49. Evin A, Dobney K, Schafberg R, Owen J, Vidarsdottir US, Larson G, Cucchi T. 2015 Phenotype and animal domestication: A study of dental variation between domestic, wild, captive, hybrid and insular Sus scrofa. *BMC Evol. Biol.* **15**, 6.

50. Kruska D, Röhrs M. 1974 Comparative--quantitative investigations on brains of feral pigs from the Galapagos Islands and of European domestic pigs. *Z. Anat. Entwicklungsgesch.* **144**, 61–73.

51. McIntosh GH, Pointon A. 1981 The Kangaroo Island strain of pig in biomedical research. *Aust. Vet. J.* **57**, 182–185.

52. Burton JA, Macdonald AA. 2008 Sus celebensis. *The IUCN Red List of Threatened Species* (doi:10.2305/IUCN.UK.2008.RLTS.T41773A10559537.en)

53. Frantz et al. 2018 Data From: Synchronous diversification of Sulawesi's iconic artiodactyls driven by recent geological events. Dryad Digital Repository. (https://doi.org/10.5061/dryad.dv322)

Data Accessibility

All datasets, including microsatellites, mitochondrial, morphometric and meta data, are available on Dryad (https://doi.org/10.5061/dryad.dv322) [53]. The mitochondrial data is also available on GeneBank (accession MH021990-MH022712).

Authors' contributions

L.A.F.F., J.B., P.G., D.S., R.H., A.C.K. A.A.M., and G.L. conceived the study. L.A.F.F. and GL wrote the paper with input from all authors. L.A.F.F., S.Y.W.O., A.R., A.M.S.N., A.E., J.B., A.H-B., A.L., G.L., P.G., D.S., E.K.I-P. analysed the data. All other authors provided samples, data and analytical tools.

Competing interests

The authors have no competing interests.

Acknowledgments

We thank Joshua Schraiber and Erik Meijaard for valuable comments. L.A.F.F., J.H, A.L., A. H-B and G.L. were supported by a European Research Council grant (ERC-2013-StG-337574-UNDEAD) and Natural Environmental Research Council grants (NE/K005243/1 and NE/K003259/1). L.A.F.F. was supported by a Junior Research Fellowship (Wolfson College, University of Oxford) and a Wellcome Trust grant (210119/Z/18/Z). P.G. S.G., J.v.d.H, C.A. and D.O. were supported by Flemish government structural funding. A.R was supported by a Marie Curie Initial Training Network (BEAN—Bridging the European and Anatolian Neolithic, GA no. 289966) awarded to M.G.T. M.G.T. is supported by a Wellcome Trust Senior Research Fellowship (GA no. 100719/Z/12/Z). B.J.E was supported by the Natural Science and Engineering Research Council of Canada. This work received additional support from the University of Edinburgh Development Trust, the Roslin Institute, the Balloch Trust and the Stichting Dierentuin Helpen (Consortium of Dutch Zoos). Additional support was also provided by The Rufford Small Grant, Royal Geographical Society, London, the Royal Zoological Society of Scotland and The University of Edinburgh Birrell-Gray Travel Award. We also thank the National Museums of Scotland for logistic support, and the Negaunee Foundation for their continued support of a curatorial preparator. We are also indebted to the Indonesian Ministry of Forestry, Jakarta (PHKA), Sulawesi's Provincial Forestry Departments (BKSDA); the Indonesian Institute of Science (LIPI); Museum of Zoology, Research Center for Biology, Cibinong (LIPI); and the project's long-standing Indonesian sponsor, Ir. Harayanto MS, Bogor Agricultural University (IPB) for sample collection/permission.

Figure legends

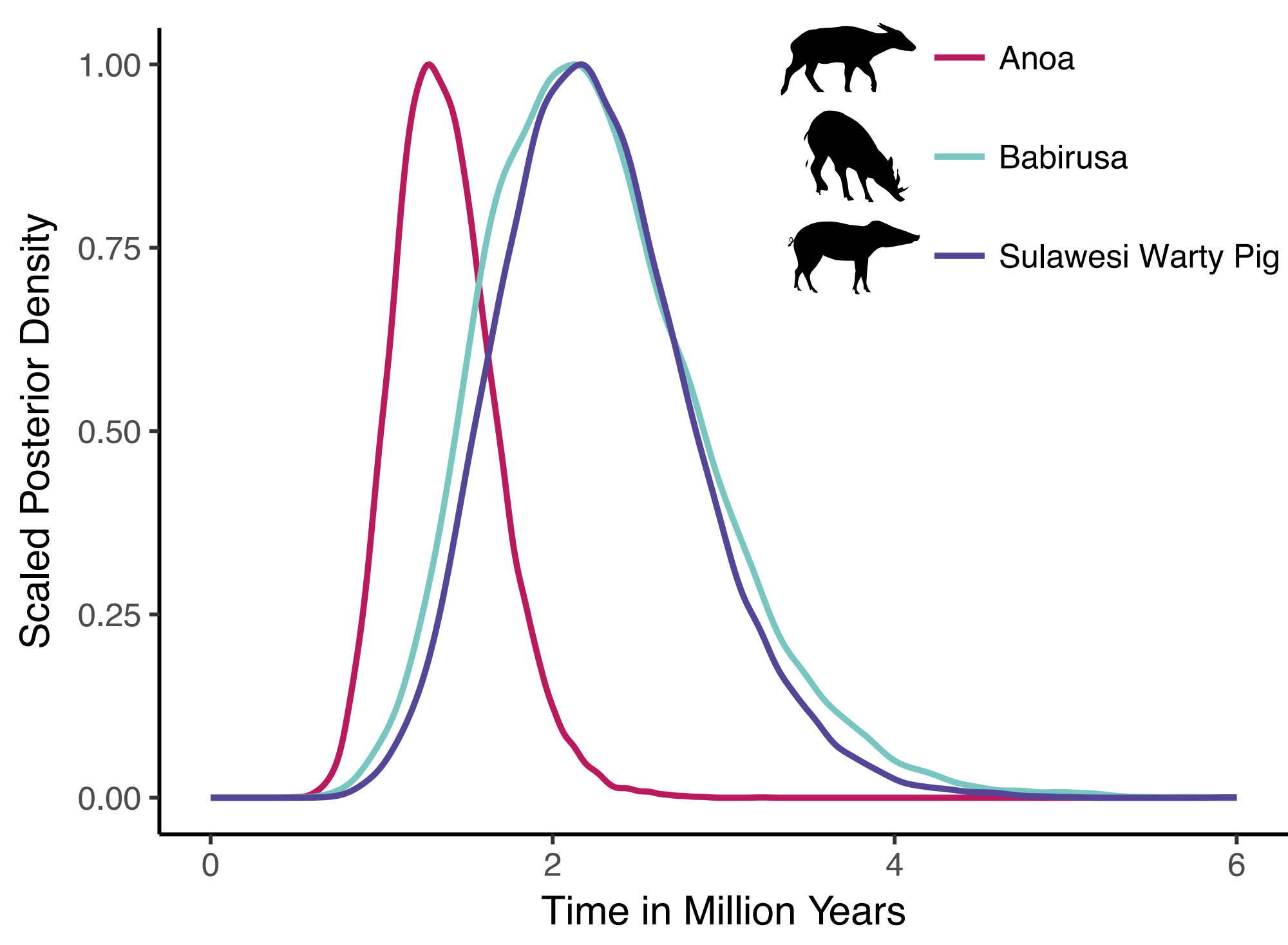
Figure 1: Time to the most recent common ancestor (TMRCA) for three mammal species on Sulawesi. Posterior densities of the TMRCA estimates for Anoa, Babirusa, and Sulawesi warty pig inferred using a Bayesian molecular clock based on mitochondrial DNA sequences.

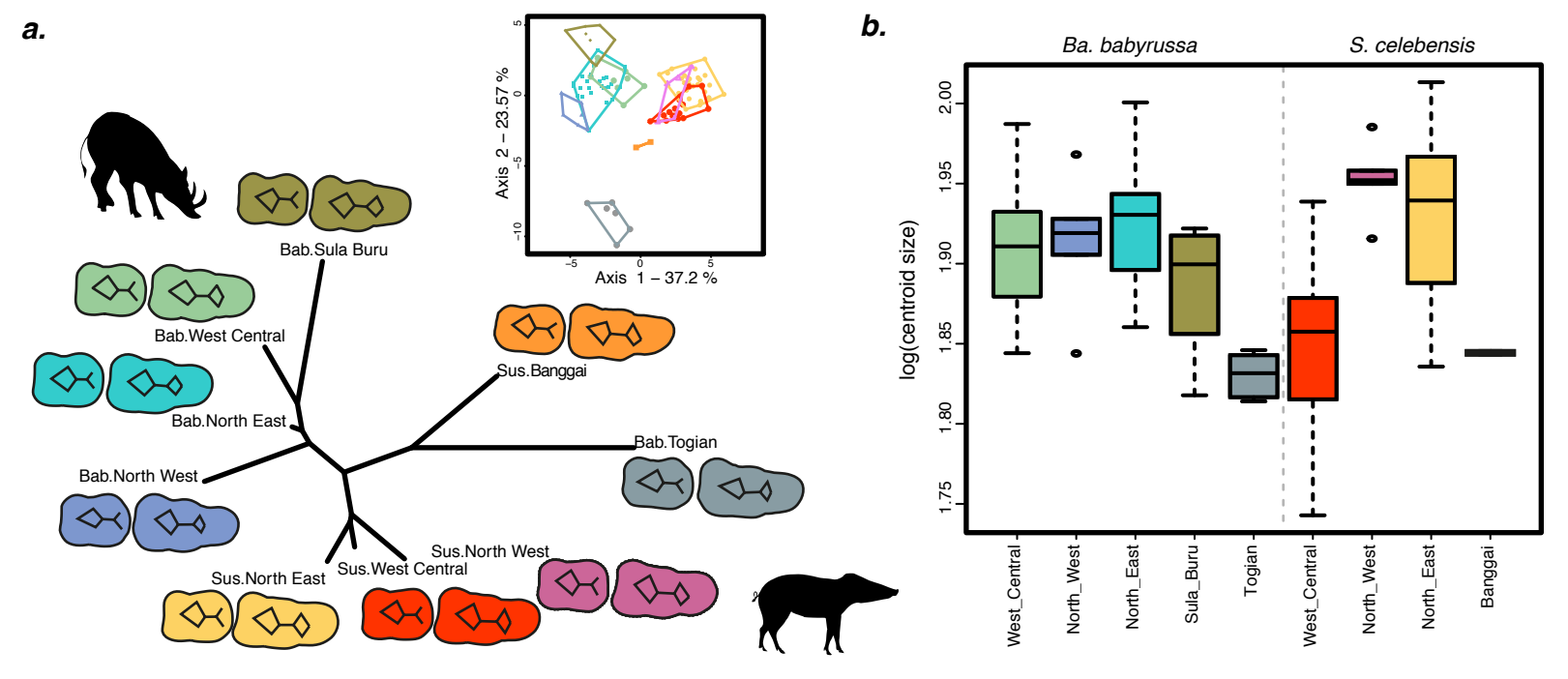
Figure 2: Population morphological variation inferred from geometric morphometric data. a. Neighbour-joining network based on Mahalanobis distances measured from second and third lower molar shapes and visualisation of population mean shape. **b.** Variation of third molar size per population (log centroid size)**.**

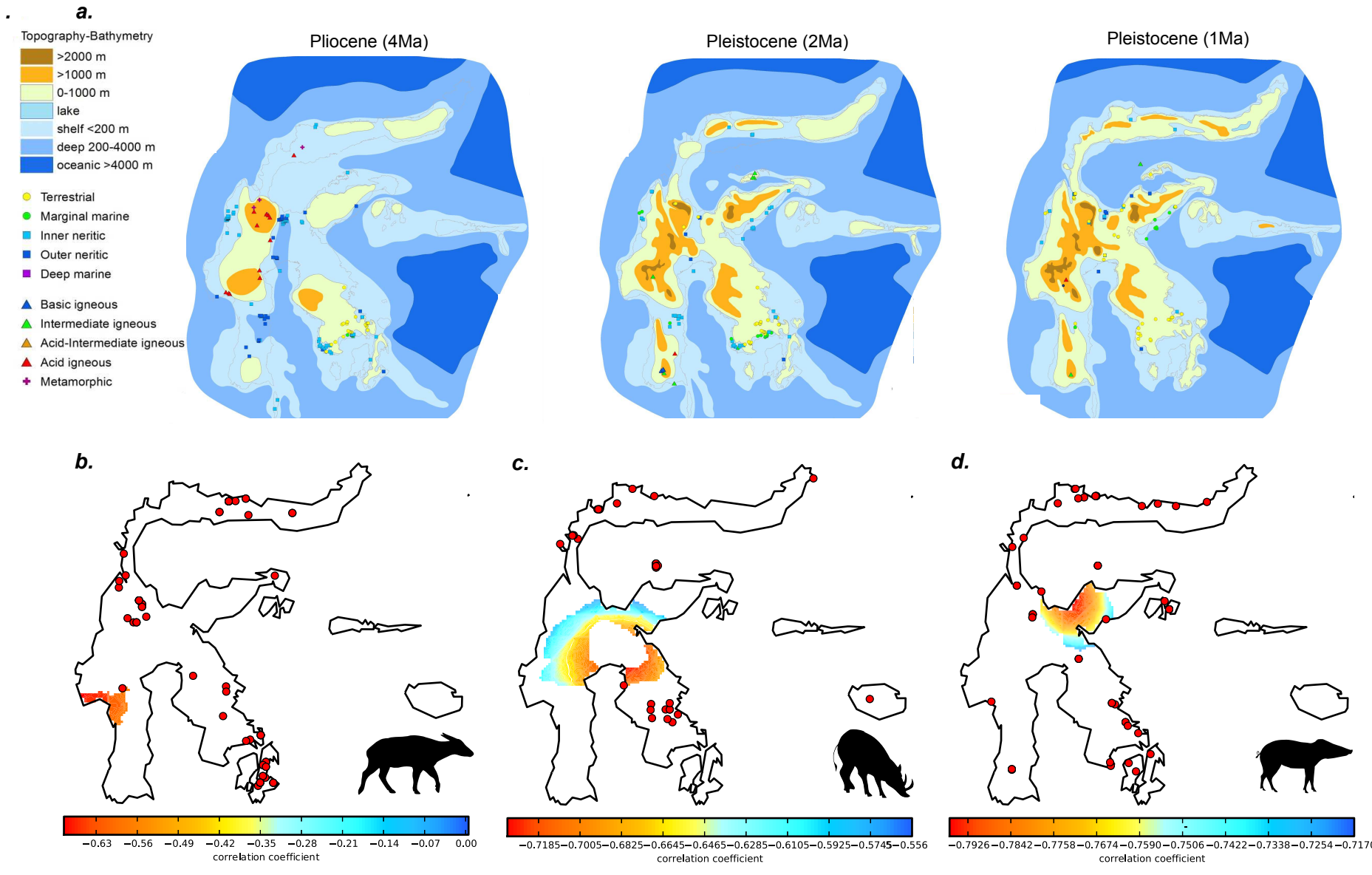
Figure 3: Geological maps of Sulawesi and the geographical origin of expansion. a. Reconstruction of Sulawesi over the last 5 Myr (adapter from [39]) and potential origin of expansion of **b.** Anoa, **c.** Babirusa, and **d.** Sulawesi warty pig. Red dots represent the location of the samples used for this analysis. Low correlation values (between distance and extrapolated genetic diversity; see *Electronic Supplementary Material*) represent most likely origin of expansion.

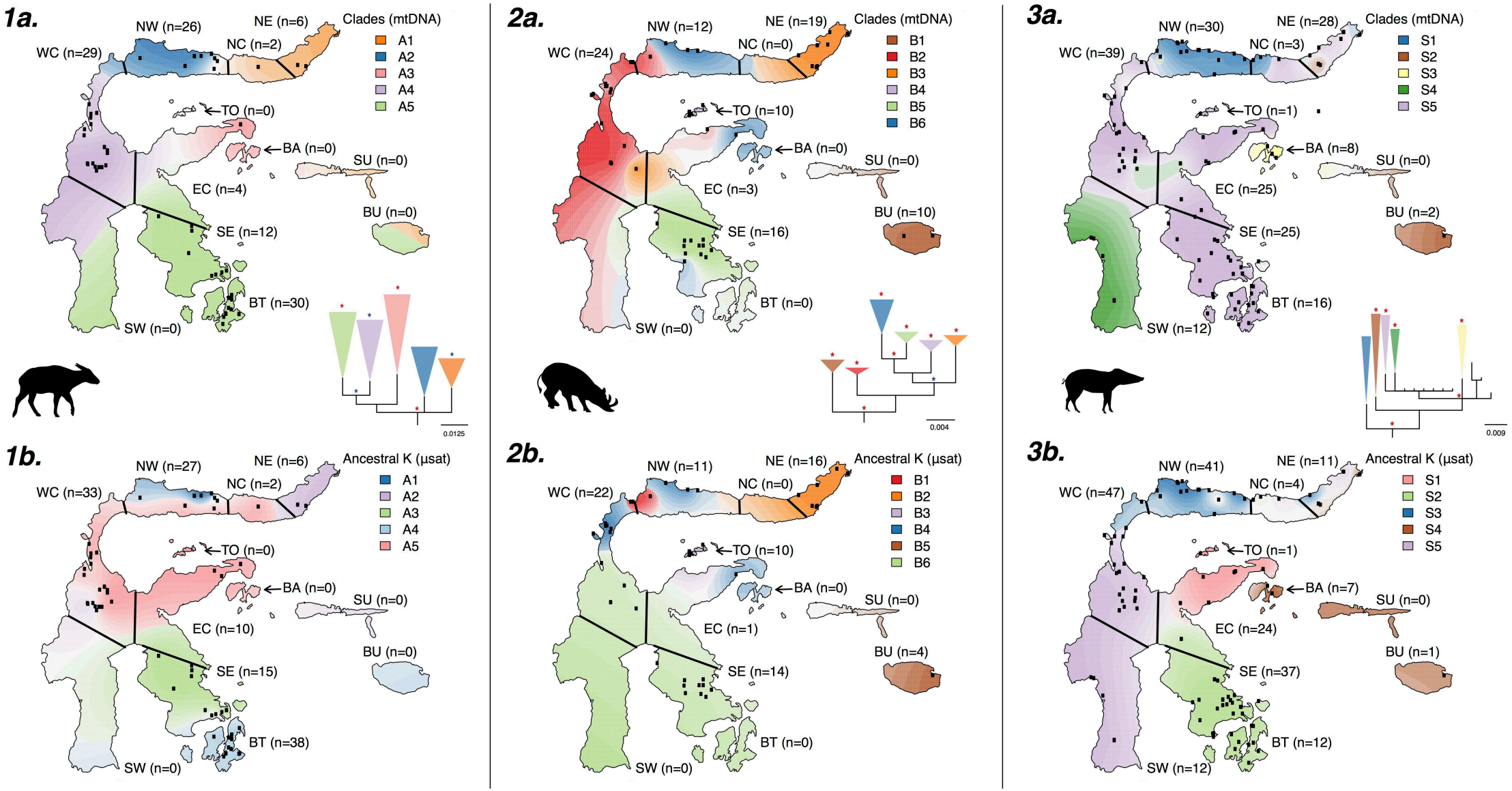
Figure 4: Population structure and geographic patterning of three mammal species on Sulawesi inferred from mitochondrial and microsatellite DNA. a. to c., A tessellated projection of sample haplogroups in each region of endemism, and phylogeny of **1.** Anoa **2.** Babirusa, and **3.** Sulawesi warty pig. Each region is labelled with the number of samples used for the projection. The projection extends over regions with no samples (e.g. the Southwest peninsula for Babirusa and Anoa) and the population membership affinities for these regions are therefore unreliable. Red and blue stars on the phylogenetic trees correspond to posterior probabilities greater than 0.9 and 0.7, respectively. **1a, 2a, 3a.** Tessellated projection of the STRUCTURE analysis, using the microsatellite data, for **2a** Anoa, **2b** Babirusa, and **2c** Sulawesi warty pig. The best *K* value for each species was used (*K*=5 for Anoa; *K*=6 for Babirusa; *K*=5 for Sulawesi warty pig; *Electronic Supplementary Material* Figure S8). NE=North East; NC=North Central; NW=North West; TO=Togian; BA=Banggai Archipelago; EC=East Central; WC=West Central; SU=Sula; BU=Buru; SE=South East; SW= South West; BT=Buton.

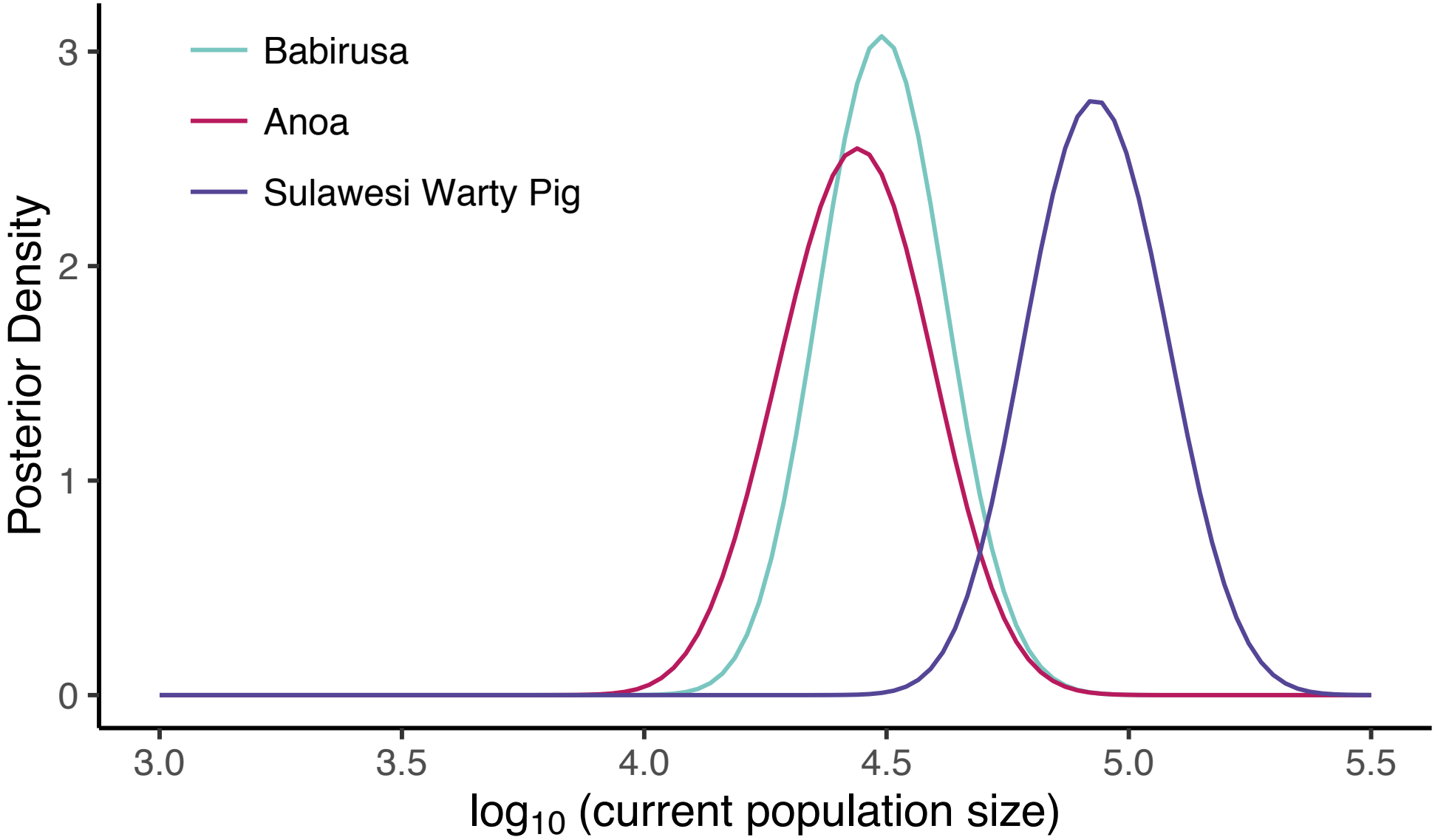
Figure 5: Posterior distribution of the current population size (Ne) of each species as inferred via approximate Bayesian computation.


Anoa
Babirusa
Sulawesi Warty Pig
Scaled Posterior Density
1.00
0.75
0.50
0.25
0.00
0
2
4
6
Time in Million Years


a.
Bab.Sula Buru
Bab.West Central
Bab.North East
Bab.North West
Sus.Banggai
Bab.Togian
Sus.North West
Sus.West Central
Sus.North East
Axis 2 – 23.57 %
Axis 1 – 37.2 %
b.
Ba. babyrussa
S. celebensis
log(centroid size)
1.75
1.80
1.85
1.90
1.95
2.00
West_Central
North_West
North_East
Sula_Buru
Togian
West_Central
North_West
North_East
Banggai


a.
Topography-Bathymetry
>2000 m
>1000 m
0-1000 m
lake
shelf <200 m
deep 200-4000 m
oceanic >4000 m
Terrestrial
Marginal marine
Inner neritic
Outer neritic
Deep marine
Basic igneous
Intermediate igneous
Acid-Intermediate igneous
Acid igneous
Metamorphic
Pliocene (4Ma)
Pleistocene (2Ma)
Pleistocene (1Ma)
b.
c.
d.
−0.63 −0.56 −0.49 −0.42 −0.35 −0.28 −0.21 −0.14 −0.07 0.00
correlation coefficient
−0.7185−0.7005−0.6825−0.6645−0.6465−0.6285−0.6105−0.5925−0.5745−0.556
correlation coefficient
−0.7926−0.7842−0.7758−0.7674−0.7590−0.7506−0.7422−0.7338−0.7254−0.7170
correlation coefficient


1a.
NW (n=26)
NE (n=6)
Clades (mtDNA)
WC (n=29)
NC (n=2)
A1
A2
A3
A4
A5
←TO (n=0)
←BA (n=0)
SU (n=0)
EC (n=4)
BU (n=0)
SE (n=12)
BT (n=30)
SW (n=0)
0.0125
1b.
NW (n=27)
NE (n=6)
Ancestral K (μsat)
WC (n=33)
NC (n=2)
A1
A2
A3
A4
A5
←TO (n=0)
←BA (n=0)
SU (n=0)
EC (n=10)
BU (n=0)
SE (n=15)
BT (n=38)
SW (n=0)
2a.
NW (n=12)
NE (n=19)
Clades (mtDNA)
WC (n=24)
NC (n=0)
B1
B2
B3
B4
B5
B6
←TO (n=10)
←BA (n=0)
SU (n=0)
EC (n=3)
BU (n=10)
SE (n=16)
BT (n=0)
SW (n=0)
0.004
2b.
NW (n=11)
NE (n=16)
Ancestral K (μsat)
WC (n=22)
NC (n=0)
B1
B2
B3
B4
B5
B6
←TO (n=10)
←BA (n=0)
SU (n=0)
EC (n=1)
BU (n=4)
SE (n=14)
BT (n=0)
SW (n=0)
3a.
NW (n=30)
NE (n=28)
Clades (mtDNA)
WC (n=39)
NC (n=3)
S1
S2
S3
S4
S5
←TO (n=1)
←BA (n=8)
SU (n=0)
EC (n=25)
BU (n=2)
SE (n=25)
BT (n=16)
SW (n=12)
0.009
3b.
NW (n=41)
NE (n=11)
Ancestral K (μsat)
WC (n=47)
NC (n=4)
S1
S2
S3
S4
S5
←TO (n=1)
←BA (n=7)
SU (n=0)
EC (n=24)
BU (n=1)
SE (n=37)
BT (n=12)
SW (n=12)


Babirusa
Anoa
Sulawesi Warty Pig
Posterior Density
0
1
2
3
3.0
3.5
4.0
4.5
5.0
5.5
log10 (current population size)

Electronic Supplementary Materials

Materials and Methods:

Sampling

We obtained DNA or morphometric samples (traditional or geometric morphometric measurements), or both, from 456 Sulawesi warty pigs (*Sus celebensis*), 520 Anoas (*Bubalus* spp.), and 313 Babirusas (*Babyrousa* spp.). Sampling on Sulawesi can be difficult due to its remoteness and to recent population declines of endemic mammals. To overcome this limitation, we targeted the extensive collections of these three species in museums, private collections, local markets, and zoos across the world. All information necessary to assess the provenance, type of specimen, and more are provided as supplementary data (Table S1).

Taxonomic notes

We sampled individuals from the geographic locations (Togian, mainland Sulawesi, and Buru/Sula) of all three Babirusa species (Table S1). For Anoas, while the majority of our samples are from specimens with no species designation (*e.g.* museum samples collected prior to the split of Anoa into two species [1]) , our data set includes individuals assigned to both lowland and highland Anoa (Table S1). Given that the goal of this study is to understand the general evolutionary history of the island (and the fact that both Anoa and Babirusa are only found in Sulawesi and the neighboring islands), we treated all the Babirusa and Anoa samples as a single taxonomic unit. The relevance of the data presented here to our understanding of species designations will be addressed in future studies.

Morphometrics

A total of 356 teeth from 227 specimens (Babirusa: 76 M2 and 89 M3; SWP: 99 M2 and 92 M3) were measured and analysed using geometric morphometric approaches in 2D. We

strictly followed protocols developed by [2,3]. Differences in shape were tested using MANOVA, whereas differences in log-transformed centroid size were tested using Wilcoxon tests and visualized using boxplots. Variation in shape was first visualized using a principal-components analysis (PCA) before between-groups variation was explored using Canonical Variate Analyses (CVA). The resemblance between groups was visualized with a neighbor-joining network calculated on the Mahalanobis distances. Manova and CVA were performed after a dimensionality reduction of the data following [3]. The variances of the two species on Sulawesi were compared using a Fligner-Killeen test based on the distance between each specimen and the mean shape (or size) of its species. M2 and M3 were analysed separately before being pooled together to produce the synthetic Figure 2a.

Genetics

DNA extraction

We extracted DNA from 520 Anoas, 251 Babirusas, and 317 SWPs. We sequenced mitochondrial cytochrome *b* (cytb) and D-loop (total length 1,394 bp) from 142 samples of Anoa, as well as partial D-loop from 213 and 230 samples of Babirusa (481 bp) and SWP (660 bp), respectively. We also typed 13 microsatellite loci for 163 samples of Anoa, 14 loci for 238 samples of SWP, and 13 loci for 182 samples of Babirusa. Genomic DNA was extracted from museum specimens, hair follicles and faeces using the DNeasy Blood and Tissue kit (Qiagen). DNA was quantified in a Nanodrop and visualized under UV light in 40 mL 1X TAE 1% agarose gels stained with SYBRsafe (Invitrogen).

For DNA extraction from bone, we grounded samples of cortical bone to powder in a Mikrodismembrator (Sartorius). We then digested bone powder overnight at 50 °C in 2 mL of buffer (0.425 M EDTA pH8, 1 mM Tris–HCl pH8, 0.05% w/v SDS, 0.33 mg/mL

Proteinase K) under constant rotation. The digested solution was concentrated to approximately 500 µL using 30 kDa molecular weight cut-off centrifugal filters (Amicon® Ultra, Millipore). We passed the concentrated solution through a silica column (QIAquick®, Qiagen) following the manufacturer's protocol, and eluted the final extract in 100 µL of TE buffer. We measured DNA concentration (Table 1) using 2 µL of extract on the Qubit® platform (Invitrogen), and stored the extracts at -20 °C.

mtDNA sequence data

From our samples of Anoa, we amplified D-loop and cytb fragments by polymerase chain reaction (PCR) using the primers described in Table S6. Both primers were designed by Dr D. Bradley (Trinity College, Dublin) to amplify the mtDNA of multiple bovine species [4,5]. Numerous samples were not sequenced due to the low quality of their DNA. Fragments were amplified by PCR using one cycle of denaturation at 96 °C for 3 min, followed by 30 cycles of: denaturation at 96 °C for 30 s, annealing at 50 °C for 20 s, and extension at 60 °C for 4 min. Both primers were run separately with an M13 tail added to the 5'-end. Sequencing was carried out using M13 universal primers and the ABI BigDye 3.1 sequencing kit (Applied Biosystems). Sequences were determined using an ABI 3700 automatic DNA capillary sequencer (Applied Biosystems), OrbixWeb™ Deamon software, 3700 DATA collection software and DATA Extractor software.

From our samples of Babirusa and *Sus celebensis*, we amplified two overlapping d-loop fragments for both species, which were amplified by PCR using primers designed by G. Larson (University of Oxford, UK)[6,7] and described in Table S6. PCR mixture was as follows: 2.5 µL x 5 Taq advanced buffer (containing 1.5 mM $MgCl_2$), 2.5 µL of each primer (10 µM), 0.5 µL 200 µM dNTPs, 0.25 µL 5 Prime Taq polymerase, 1 µL DNA (50–100 ng) adjusted to a final volume of 25 µL with ddH_2O. Fragments were amplified using one cycle of denaturation at 94 °C for 1 min 30 s followed by 40 cycles of: denaturation at 94 °C for

45 s, annealing at 53 °C for 45 s, extension at 72 °C for 1 min 30 s, followed by a final extension at 72 °C for 10 min. Each fragment of either marker was subjected to bidirectional sequencing using the ABI BigDye 3.1 sequencing kit (Applied Biosystems). Sequences were generated using an ABI 3130 DNA capillary sequencer (Applied Biosystems).

Microsatellite data

Anoa samples were genotyped for 13 bovine microsatellite loci using primers previously designed for cattle *Bos taurus* (with the forward primer fluorescently labelled): BM1818, CSRM60, ETH152, HAUT24, HAUT27, HEL13, ILSTS5, INRA35, INRA37, MM12, SPS115, TGLA126, and TGLA227. These loci were recommended by the Food and Agriculture Organization [8] for use in genetic diversity studies and were selected at the Roslin Institute (Edinburgh, UK). More details and the primer sequences are available in Table S6.

For some samples, PCR were done as simplex reactions in 10 µL final volume containing 1 µL 10X PCR buffer, 0.3 µL of 50 µM $MgCl_2$, 1 µL of each primer (10 µM), 1 µL of dNTPs (10 µM), 0.1 µL Platinum Taq polymerase, 4.6 µL of ddH_2O and 1 µL DNA (50–100 ng). Simplex PCR conditions were: initial denaturation at 94 °C for 3 min, followed by 30 cycles of: denaturation at 94 °C, annealing at 55–65 °C (depending on the marker) for 45 s and extension at 72 °C for 45 s, with a final extension of 72 °C for 3 min. For other samples, PCRs were done as multiplex reactions by pooling six or seven microsatellite primer pairs using the Type-It Microsatellite kit (Qiagen). Multiplex reactions were done in a final volume of 10 µL containing 5 µL 2X Type-It Master Mix, 1 µL 10X primer mix, 1 µL Q-solution, 1 µL ddH_2O and 2 µL DNA (50–100 ng). Multiplex PCR conditions followed the manufacturer's instructions (Qiagen). DNA from *Bos taurus* was used as a positive control.

Negative controls (without DNA) were included in all reactions. The PCR products were analysed using an ABI 373 (Applied Biosystems) DNA fragment analyser. Results were scored with the programs GENESCAN 3.0, GENOTYPER 2.5 or PEAK SCANNER 2.0 (Life Technologies).

For samples of *Sus celebensis*, PCR was performed in an Eppendorf Mastercycler® gradient apparatus. In general, the PCR profile was as follows: the 10 µL reaction mixture consisted of 1 µL DNA (about 50–100 ng), 1 x 5 Prime Taq advanced buffer (containing 1.5 mM $MgCl_2$), 1 µL of M13F (1 µM), 1 µL of each primer (10 µM) (0.5 µL for S0214 and S0149), 0.2 µL 200 µM dNTPs, 0.05 µL 5 Prime Taq DNA polymerase (0.1 µL for S0214 and S0149), 1 µL DNA (50–100 ng) adjusted to a final volume of 10 µL with ddH_2O. The thermal cycling, preceded by 5 min at 94 °C and followed by 5 min at 72 °C, consisted of 30 cycles (32 for S0386 and 35 for S0026) of 94 °C for 1 min, an optimal annealing temperature for 1 min (Table S6), and 72 °C for 1 min. PCR products were visualized on a 1.5% agarose gel (Acros organics) with GelRed Nucleic Acid Gel Stain (Biotium) in order to check the amplification.

Fragment analysis was performed on an ABI 310 (Life Technologies). For all markers, we used the M13 method to visualize the PCR products. To do so we added a M13 Forward (M13F; 5'-CACGACGTTGTAAAACGAC-3') tag to the 5' end of each forward primer. PCR mix contained 0.1 µM of this tag-labelled primer and 1 µM of both reverse primer and M13F labelled primer (0.05µM of tag-labelled primer and 0.5µM of reverse primer; and M13F labelled primer for markers S0149 and S0214). Data were interpreted and allele sizes determined using GeneMapper 4.0 software (Life Technologies).

For samples from Babirusa, PCRs were performed in an Eppendorf Mastercycler® gradient apparatus. In general, the PCR profile was as follows: the 10 µL reaction mixture consisted of 1 µL DNA (about 50–100 ng), 1 x Eppendorf Taq buffer containing 1.5 mM $Mg(Oac)_2$, 1 µM of each primer (0.5 µM for S0214 and S0149), 200 µM dNTPs (Eppendorf) and 0.25 U Taq DNA polymerase (0.5 U for S0214 and S0149). The thermal cycling, preceded by 5 min at 94 °C and followed by 5 min at 72 °C, consisted of 30 cycles (32 for S0386 and 35 for S0026) of 94 °C for 1 min, an optimal annealing temperature for 1 min (see Table S6), and 72 °C for 1 min. PCR products were visualized on a 1.5% agarose gel (Acros organics) with ethidium bromide (Merck) in order to check the amplification.

Fragment analysis was performed on an A.L.F. express DNA Sequencer (Pharmacia Biotech). For markers S0149 and S0228, we used the M13 method (Boutin-Ganache et al 2001) to visualize the PCR products. Hence, an M13 Forward (5'-CACGACGTTGTAAAACGAC-3') tag was added to the 5' end of each forward primer and the PCR mix contained 0.1 µM of this tag-labelled primer and 1 µM of the reverse primer as well as of the M13F-cy5 labelled primer (or 0.05 µM of the tag-labelled primer and 0.5 µM of the reverse and M13F-cy5 labelled primer in case of marker S0149). Data were interpreted and allele sizes determined using Genetools from SynGene and Allelelocator 1.03 software (Pharmacia Biotech). All primers are available in Table S6.

Phylogenetic analyses of mitochondrial DNA

A phylogenetic tree was inferred for each species, using a Bayesian approach implemented in MrBayes v3.2.5 [9](Figure S4; Figure S5; Figure S6). To estimate the position of the root, we included a sequence from *Phacochoerus africanus* (accession: AJ314533) for the analysis of Babirusa and SWP, and from *Bos taurus* (accession:

EU177842) for the analysis of Anoa. The HKY+G substitution model was selected, for each data-set, based on Bayes factors (marginal likelihood computed via stepping-stone sampling) of JC, HKY+G and GTR+G, with and without invariable sites. To estimate the posterior distribution of various parameters, we used Markov chain Monte Carlo sampling with 4 chains (comprising 3 heated chains and 1 cold chain) of 10,000,000 steps each (with samples drawn every 1000 steps). The first 25% of samples were discarded as burn-in. We carried out 4 independent MCMC analyses and combined the samples from the posterior. Convergence was assessed by ensuring that average standard deviation of split frequencies was below 0.01 and that the potential scale reduction factor was close to 1 for all parameters.

For each species we defined haplogroups based on highly supported clades. For each geographic region, the proportion of each haplogroup was plotted on a map using the R package “maps”. For each sample, haplogroup membership was transposed to create an ancestry matrix. All samples lacking precise geographic coordinates were removed. The ancestry matrix was then plotted onto a map with a tessellated projection, using the R package “tess3r” [10–12]. We then divided Sulawesi and nearby islands into 11 regions based on previous work on amphibians and primates that defined areas of endemism on the island [13–15]. We assessed the significance of the difference in haplogroup frequency in each area of endemism using Pearson's chi-squared test, *p*-values were computed using 2000 simulation replicates, as implemented in R.

To infer the evolutionary timescales of the three species, we performed a Bayesian phylogenetic analysis using a molecular clock in BEAST v1.8.4 [16]. First, we analysed a mtDNA combined data set comprising the sequences of *Babyrousa* spp., *Sus celebensis* and relatives (*S. cebifrons*, *S. philippensis*, *Hylochoerus meinertzhageni*, *Potamochoerus*

porcus, *Potamochoerus larvatus*, *Phacochoerus aethiopicus*, and *Phacochoerus africanus*). This data set comprised 700 aligned nucleotides from 243 samples. To calibrate the molecular clock, we used a normal calibration prior for the age of African suids (mean 10.5 My, standard deviation 2.551 My), based on the estimate from a combined nuclear and mitochondrial data set by [17].

We then analysed the mtDNA sequences of *Bubalus* spp. and related bovids (*Bison bison*, *Bison bonasus*, *Syncerus caffer*, *Bos taurus*, *Bos gaurus*, *Bos frontalis*, and *Bos grunniens*). This data set comprised 726 aligned nucleotides from 170 samples. We used a normal calibration prior for the age of the root (mean 8.8 My, standard deviation 1.02041 My), based on a fossil calibration used by [18]. Given the use of relatively deep calibrations in both analyses, the date estimates should be regarded as being conservatively old because our approach is likely to produce underestimates of the substitution rates [19].

The Bayesian information criterion was used to select the HKY+G model as the best-fitting substitution model for both data sets, after excluding models allowing a proportion of invariable sites. For each data set we compared two models of rate variation: the strict clock and the uncorrelated lognormal relaxed clock [20]. We also compared three tree priors: constant-size coalescent prior, Bayesian skyline coalescent prior, and birth-death speciation prior. For each combination of clock model and tree prior, the marginal likelihood was estimated using path sampling with 25 power posteriors [21]. Samples were drawn every 2,000 steps from a total of 2,000,000 MCMC steps per power posterior.

Posterior distributions of all parameters, including the tree, were estimated by MCMC sampling, with samples drawn every 5000 steps over a total of 50,000,000 MCMC steps.

To ensure convergence, each analysis was run in duplicate and the samples were compared and combined. Sufficient sampling was confirmed by examining the effective sample sizes of parameters. For both data sets, the strict clock and Bayesian skyline tree prior yielded the highest marginal likelihood (Table S7).

Analyses of microsatellite data

For each species, we used STRUCTURE v2.3.4 [22] to infer population structuring. The maximum number of populations (*K*) was set to 12 (the total number of region defined on Sulawesi). For each species, we ran 10 independent MCMC analyses, each with 1,000,000 steps, discarding a burn-in of 50,000 steps. We computed ΔK (Figure S8) to infer the best-fitting *K* value using structure Harvester [23]. Independent runs were merged using *CLUMPP* with *M*=2 [24]. For all samples with precise geographic coordinates, results were plotted onto a map with a tessellated projection, using the R package "tess3r" [10–12]. Results were also plotted on a map using the R package "maps" in each region of endemism (see above). To limit the possibility of provenance uncertainty, we excluded all samples that were from zoos or from unknown locations from this analysis (Table S1).

We used the package hierfstat v0.04 [25] in R to compute Weir and Cockerham's *Fst* [26]. Analyses of molecular variance (AMOVA)[27] were also performed in R using the package poppr v2.3.0 [28] and ade4 v1.7 [29] using populations as defined in Figure 4. We built neighbour-joining trees based on pairwise proportions of shared alleles [30](POSA; Figure S12) using PHYLIP [31]. For *Babyrousa* spp. and SWP we also computed average square distance (ASD) [32] between every pair of samples at 13 microsatellite loci (shared between SWP and Babirusa) in order to estimate the relative TMRCAs of these species [33]. Both ASD and POSA were computed using Microsatellite Analyser v3.13[34].

Geographical origins of population expansions

To infer the location of origin of population expansion for each species, we employed a spatially explicit discriminative modelling approach in which we assume a monotonic decline in diversity with distance from origin of a range expansion. A spatial grid of latitude and longitude values covering the geographic space of Sulawesi, of resolution 0.05 by 0.05 degrees, was explored using a flat kernel of radius 500 km for SWP and Babirusa and 350 km for Anoa. If at any location in the grid we found within the kernel at least 5 sampled individuals for SWP, or 3 sampled individuals for Babirusa and Anoa, the local diversity was calculated using ASD and recorded for that grid location. The grid was then re-explored with each latitude/longitude location treated as a potential origin location, and we recorded the correlation between geographic distance to the accepted kernels and local diversity at those kernels. This provided a grid of correlation values, which was then interpolated and visualized on a map.

Regions with the highest negative correlations were considered the best hypothesized origin locations. To quantify statistical support for inferred origin locations, the data were permuted among sample sites 1000 times, and for each permuted data set the above analysis was repeated. Following this, we plotted only the grid locations where the negative correlation between geographic distance and genetic diversity was more extreme than 99% (98% for Anoa) of those obtained from the permuted data.

Approximate Bayesian computation

For each species, we used both mtDNA and microsatellite data to evaluate the fit of four different models (Figure S11) and to obtain a posterior distribution of the parameters under the best-fitting model. We compared the fit of models with constant population size (Figure S11a), population expansion (Figure S11b), a bottleneck (Figure 10c), and a bottleneck

following an expansion (Figure 10d). The rationale behind these models is to test whether these species have undergone a population expansion due to the uplift of Sulawesi (see main text) and/or if they have undergone a bottleneck due to recent human activities. The prior distributions used for the simulations are summarized in Table S4.

We calculated multiple summary statistics for each data set using arlsumstat [35]. For the mtDNA data, we computed the number of segregating haplotypes *K*, the number of segregating sites *S*, Tajima's *D [36]*, Fu's *FS [37]*, and the average pairwise difference π. For the microsatellite data, we computed the total number of alleles *K*, the range of the allele size *R*, the expected heterozygosity *H* and the Garza–Williamson statistic *GW [38]*. We ensured that the observed summary statistics fell well within the distribution of simulated summary statistics (Figure S13-15).

For model-testing purposes, we performed 200,000 simulations per model using fastsimcoal2 [39]. We chose a set of informative summary statistics with a partial least-squares discriminant analysis as in [40,41] using the *plsda* function in R [42]. We compared all models (computing marginal likelihood and posterior probability) simultaneously using a standard ABC generalized linear model (GLM) approach as implemented in ABCtoolbox [43]. We also computed the average Root Mean Square Error (RMSE) for each parameter using pseudo-observed data to assess our power to infer each parameter in the model (see Table S4).

To estimate parameter values, we ran a total of 2,000,000 simulations under the best-fitting model for each species. We extracted five partial least square (PLS) components from the summary statistics in the observed and simulated data [44]. We retained a total of 10,000 simulations closest to the observed data and applied a standard ABC-GLM [45].

Supplementary Figures:

Figure S1. Venn diagram representing the number of individuals and the overlap between the various databases generated for this project. a. Anoa **b.** Babirusa **c.** *Sus celebensis.*

Figure S2: Molecular clock results for suids alignment

Figure S3: Molecular clock results for bovids alignment

Figure S4: Bayesian phylogeny inferred from mtDNA from *Sus celebensis*. Support values represent posterior probabilities, S1-5 label represent haplogroups plotted in Figure 1.

Figure S5: Bayesian phylogeny based on mtDNA from Babirusa. Support values represent posterior probabilities; B1-6 labels represent haplogroups plotted in Figure 1.

Figure S6: Bayesian phylogeny based on mtDNA from Anoa. Support values represent posterior probabilities; A1-5 labels represents haplogroups plotted in Figure 1.

Figure S7: Tectonic reconstruction of Sulawesi over the last 8My in 1My increments adapted from [46]

Figure S8: ΔK values for each species (best number of clusters in the microsatellite data). a. Anoa **b.** Babirusa **c.** Sulawesi warty pig.

Figure S9: Neighbour-joining trees based on Fst. a. Anoa **b.** Babirusa **c.** Sulawesi warty pig.

Figure S10: Results of the STRUCTURE analysis for *K*=2 to *K*=6. a. Anoa **b.** Babirusa **c.** Sulawesi warty pig.

Figure S11: Various models tested using approximate Bayesian computation. a. Constant population size (Model 1). **b.** Population expansion (Model 2). **c.** Population bottleneck (Model 3). **d.** Population expansion followed by a bottleneck (Model 4).

Figure S12: Neighbour-joining tree based on pairwise proportion of shared alleles using the microsatellite data. a. Anoa **b.** Babirusa **c.** Sulawesi warty pig.

Figure S13 Observed (red vertical line) and simulated (histogram) of all summary statistics used in the approximate Bayesian computation analysis (Anoa).

Figure S14 Observed (red vertical line) and simulated (histogram) of all summary statistics used in the approximate Bayesian computation analysis (Babirusa).

Figure S15 Observed (red vertical line) and simulated (histogram) of all summary statistics used in the approximate Bayesian computation analysis (SWP).

Figure S16: Population structure of each species inferred from mtDNA, microsatellites. a. to c., Proportion of haplogroups in each region of endemism and

phylogeny of Anoa (**a.**), Babirusa (**b.**) and Sulawesi warty pig (**c.**). Numbers in pie charts represent the sample size in a given region. **d. to f.,** Result of the STRUCture analysis using the microsatellite data plotted on the map and as a bar chart (Figure S10) for Anoa (**d.**), Babirusa (**e.**) and SWP (**f.**). The best *K* value for each species was used (*K*=5 for Anoa; *K*=6 for Babirusa; *K*=5 for SWP). NE=North East; NC=North Central; NW=North West; TO=Togian; BA=Banggai Archipelago; EC=East Central; WC=West Central; SU=Sula; BU=Buru; S=Sula or Buru; SE=South East; SW= South West; BT=Buton.

Supplementary Tables:

Table S1: Table containing sample information for all three species – available at https://doi.org/10.5061/dryad.dv322

Table S2: Pairwise Wilcoxon tests for the lower M3 (upper part) and lower M2 (lower part), for the lower M3 (upper part) and lower M2 (lower part).

Table S3: **Support for various models obtained from the ABC analysis.** Each models tested (1-4) are displayed in Figure S11. Obs. P-value= observed fraction of the retained simulation (2,000) with a marginal likelihood value (marginal *lnL*) smaller than the observed data. Posterior *P.* = Posterior probability of the model.

Table S4: Characteristics of the prior and posterior distribution of parameters estimated via approximate Bayesian computation. All priors are uniformly distributed. The average root mean square error (RMSE) of the mode of each parameter was computed using 1,000 pseudo-observed data sets. Values close to 1 and 0 indicates little and large power, respectively. 95CI represents the 95% credibility interval. See Figure S11 for further information about the parameters.

Table S5: Results of the AMOVA based on microsatellite data.

Table S6: List of all primers used in this study

Table S7: Marginal likelihood of molecular clock analyses under constant-size coalescent prior, Bayesian skyline coalescent prior, and birth-death speciation prior.

References:

1. Groves CP. 1969 Systematics of the anoa (Mammalia, Bovidae). *Beaufortia* **17**, 1–12.

2. Cucchi T, Hulme-Beaman A, Yuan J, Dobney K. 2011 Early Neolithic pig domestication at Jiahu, Henan Province, China: clues from molar shape analyses using geometric morphometric approaches. *J. Archaeol. Sci.* **38**, 11–22.

3. Evin A, Cucchi T, Cardini A, Strand Vidarsdottir U, Larson G, Dobney K. 2013 The long and winding road: identifying pig domestication through molar size and shape. *J. Archaeol. Sci.* **40**, 735–743.

4. Cymbron T, Loftus RT, Malheiro MI, Bradley DG. 1999 Mitochondrial sequence variation suggests an African influence in Portuguese cattle. *Proc. Biol. Sci.* **266**, 597–603.

5. Schreiber A, Seibold I, Nötzold G, Wink M. 1999 Cytochrome b gene haplotypes characterize chromosomal lineages of anoa, the Sulawesi dwarf buffalo (Bovidae: Bubalus sp.). *J. Hered.* **90**, 165–176.

6. Larson G *et al.* 2005 Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307**, 1618–1621.

7. Larson G *et al.* 2007 Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 15276–15281.

8. Bradley DG, Fries R, Bumstead N, Nicholas FW, Cothran EG, Ollivier L, Crawford AM. 2004 Secondary Guidelines for Development of National Farm Animal Genetic Resources Management Plans. *Food and Agricultural Organization of United Nations (FAO), Roma, Italy*

9. Ronquist F *et al.* 2012 MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst. Biol.* , sys029–.

10. Caye K, Jay F, Michel O, Francois O. 2016 Fast Inference of Individual Admixture Coefficients Using Geographic Data. *bioRxiv* , 080291.

11. Martins H, Caye K, Luu K, Blum MGB, Francois O. 2016 Identifying outlier loci in admixed and in continuous populations using ancestral population differentiation statistics. *bioRxiv* , 054585.

12. Caye K, Deist TM, Martins H, Michel O, François O. 2016 TESS3: fast inference of spatial population structure and genome scans for selection. *Mol. Ecol. Resour.* **16**, 540–548.

13. Evans BJ, Supriatna J, Andayani N, Setiadi MI, Cannatella DC, Melnick DJ. 2003 Monkeys and toads define areas of endemism on Sulawesi. *Evolution* **57**, 1436–1443.

14. Evans BJ, Supriatna J, Andayani N, Melnick DJ. 2003 Diversification of Sulawesi macaque monkeys: decoupled evolution of mitochondrial and autosomal DNA. *Evolution* **57**, 1931–1946.

15. Merker S, Driller C, Perwitasari-Farajallah D, Pamungkas J, Zischler H. 2009 Elucidating geological and biological processes underlying the diversification of Sulawesi tarsiers. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 8459–8464.

16. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012 Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973.

17. Gongora J *et al.* 2011 Rethinking the evolution of extant sub-Saharan African suids (Suidae, Artiodactyla). *Zool. Scr.* **40**, 327–335.

18. Bibi F. 2013 A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. *BMC Evol. Biol.* **13**, 166.

19. Ho SYW, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigo AG, Cooper A. 2011 Time-dependent rates of molecular evolution. *Mol. Ecol.* **20**, 3087–3101.

20. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006 Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88.

21. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012 Improving the Accuracy of Demographic and Molecular Clock Model Comparison While Accommodating Phylogenetic Uncertainty. *Mol. Biol. Evol.* **29**, 2157–2167.

22. Pritchard JK, Stephens M, Donnelly P. 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

23. Earl DA, vonHoldt BM. 2011 STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361.

24. Jakobsson M, Rosenberg NA. 2007 CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801–1806.

25. Goudet J. 2005 Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Resour.* **5**, 184–186.

26. Weir BS, Cockerham CC. 1984 Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358.

27. Excoffier L, Smouse PE, Quattro JM. 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**.

28. Kamvar ZN, Tabima JF, Grünwald NJ. 2014 *Poppr* : an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2**, e281.

29. Dray S, Dufour A-B. 2007 The **ade4** Package: Implementing the Duality Diagram for Ecologists. *J. Stat. Softw.* **22**, 1–20.

30. Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL. 1994 High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455–457.

31. Felsenstein J. 1989 PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 163–166.

32. Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW. 1995 An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**, 463–471.

33. Sun JX, Mullikin JC, Patterson N, Reich DE. 2009 Microsatellites are molecular clocks that support accurate inferences about history. *Mol. Biol. Evol.* **26**, 1017–1027.

34. Dieringer D, Schlötterer C. 2003 microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Mol. Ecol. Notes* **3**, 167–169.

35. Excoffier L, Lischer HEL. 2010 Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567.

36. Tajima F. 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.

37. Fu YX. 1997 Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchhiking and Background Selection. *Genetics* **147**, 915–925.

38. Garza JC, Williamson EG. 2001 Detection of reduction in population size using data from microsatellite loci. *Mol. Ecol.* **10**, 305–318.

39. Excoffier L, Foll M. 2011 fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* **27**, 1332–1334.

40. Peter BM, Huerta-Sanchez E, Nielsen R. 2012 Distinguishing between selective sweeps from standing variation and from a de novo mutation. *PLoS Genet.* **8**, e1003011.

41. Frantz LAF *et al.* 2015 Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat. Genet.* **47**, 1141–1148.

42. Lê Cao K-A, González I, Déjean S. 2009 integrOmics: an R package to unravel relationships between two omics datasets. *Bioinformatics* **25**, 2855–2856.

43. Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L. 2010 ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics* **11**, 116.

44. Wegmann D, Leuenberger C, Excoffier L. 2009 Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics* **182**, 1207–1218.

45. Leuenberger C, Wegmann D. 2010 Bayesian computation and model selection without likelihoods. *Genetics* **184**, 243–252.

46. Abang Mansyursyah Surya Nugraha and Robert Hall. In press. Late Cenozoic palaeogeography of Sulawesi, Indonesia. *Palaeogeogr. Palaeoclimatol. Palaeoecol.*

Table S2: Pairwise Wilcoxon tests for the lower M3 (upper part) and lower M2 (lower part), for the lower M3 (upper part) and lower M2 (lower part).

	Bab.West_Central	Bab.North_West	Bab.North_East	Bab.Sula_Buru	Bab.Togian	Sus.West_Central	Sus.North_West	Sus.North_East	Sus.Banggai
Bab.West_Centra	-	0.950	0.428	0.622	0.950	0.499	0.130	0.347	0.435
Bab.North_West	0.950	-	0.664	0.429	0.699	0.132	0.142	0.420	0.429
Bab.North_East	0.332	0.634	-	0.202	0.520	0.004	0.104	0.633	0.598
Bab.Sula_Buru	0.354	0.247	0.048	-	0.931	0.511	0.019	0.206	0.151
Bab.Togian	0.001	0.004	0.000	0.052	-	0.098	0.059	0.420	0.247
Sus.West_Centra	0.001	0.003	0.000	0.087	0.508	-	0.003	0.006	0.046
Sus.North_West	0.798	0.852	1.000	0.435	0.020	0.007	-	0.261	0.435
Sus.North_East	0.224	0.451	0.363	0.026	0.000	0.000	0.491	-	0.931
Sus.Banggai	0.524	0.662	0.105	0.841	0.017	0.077	0.354	0.068	-

Table S3: Support for various models obtained from the ABC analysis.

		Obs. *P*-value	Marginal *lnL*	Bayes Factor	Posterior *P.*
Bubalus spp.	Model 1	0	6.78E-08	4.04E-05	4.04E-05
	Model 2	0	1.00E-08	5.97E-06	5.97E-06
	Model 3	0.379	0.000365477	0.278348	0.21774
	Model 4	0.8	0.00131294	**3.59165**	**0.782213**
Babyroussa spp.	Model 1	0	8.89E-16	8.43E-13	8.43E-13
	Model 2	0	2.40E-16	2.28E-13	2.28E-13
	Model 3	0.406	0.00033359	0.462939	0.316444
	Model 4	0.673	0.000720592	**2.16011**	**0.683556**
S. celebensis	Model 1	0	3.01E-09	3.87E-05	3.87E-05
	Model 2	0	4.78E-09	6.15E-05	6.15E-05
	Model 3	0.026	1.25E-05	0.190926	0.160317
	Model 4	0.087	6.53E-05	**5.23374**	**0.839583**

Table S4: Characteristics of the prior and posterior distribution of parameters estimated via approximate Bayesian computation.

	parameter	prior_min	prior_max	RMSE	mode	HPDI-95- lower	HPDI-95- upper
B. depressicornis	N	3	5.5	0.3455	4.4394	4.13611	4.73285
	N_a/N_b	-3	0	0.9441	-1.39394	-2.98221	-0.160913
	N_b/N	0	2	0.9488	1.23232	1.02401	1.93015
	T_g	130000	440000	0.9791	233334	140986	424006
	T_b	1	15000	0.887	11970	2896	14671
B. babirussa	N	3	5.5	0.3234	4.4899	4.2436	4.74727
	N_a/N_b	-3	0	0.9774	-1.87879	-2.89784	-0.17991
	N_b/N	0	2	0.9084	1.29293	1.03074	1.93997
	T_g	330000	940000	0.9909	570303	352200	910694
	T_b	1	15000	0.8978	13485	5370	14832
S. celebensis	N	3	5.5	0.3098	4.91919	4.66545	5.2083
	N_a/N_b	-3	0	0.9795	-2.06061	-2.90735	-0.188233
	N_b/N	0	2	0.9171	1.23232	1.02349	1.92281
	T_g	330000	940000	0.995	521010	349250	904971
	T_b	1	15000	0.8942	11212	3016	14597

Table S5: Results of the AMOVA based on microsatellite data.

AMOVA Bubalus spp.

	Sigma	%
Variations Between Population	0.40	17.31
Variations Between samples Within Population	0.59	25.44
Variations Within samples	1.32	57.26
Total variations	2.31	100.00

AMOVA Babyroussa spp.

	Sigma	%
Variations Between Population	1.04	27.70
Variations Between samples Within Population	0.13	3.34
Variations Within samples	2.60	68.96
Total variations	3.77	100.00

AMOVA *S. celebensis*

	Sigma	%
Variations Between Population	0.19	4.88
Variations Between samples Within Population	0.48	12.33
Variations Within samples	3.24	82.79
Total variations	3.92	100.00

Table S6: Primers for each species

Anoa Microsatellite

Locus	Forward primer	Reverse Primer
TGLA227	CGAATTCCAAATCTGTTAATTTGCT	ACAGACAGAAACTCAATGAAAGCA
CSRM60	AAGATGTGATCCAAGAGAGAGGCA	AGGACCAGATCGTGAAAGGCATAG
TGLA126	CTAATTTAGAATGAGAGAGGCTTCT	TTGGTCTCTATTCTCTGAATATTCC
INRA037	GATCCTGCTTATATTTAACCAC	AAAATTCCATGGAGAGAGAAAC
INRA035	ATCCTTTGCAGCCTCCACATTG	TTGTGCTTTATGACACTATCCG
HEL13	AAGGACTTGAGATAAGGAG	CCATCTACCTCCATCTTAAC
MM 12	CAAGACAGGTGTTTCAATCT	ATCGACTCTGGGGATGATGT
HAUT24	CTCTCTGCCTTTGTCCCTGT	AATACACTTTAGGAGAAAAATA
HAUT27	TTTTATGTTCATTTTTTGACTGG	AACTGCTGAAATCTCCATCTTA
ILSTS5	GGAAGCAATGAAATCTATAGCC	TGTTCTGTGAGTTTGTAAAGC
ETH 152	AGGGAGGGTCACCTCTGC	CTTGTACTCGTAGGGCAGGC
SPS 115	AAAGTGACACAACAGCTTCTCCAG	AACGAGTGTCCTAGTTTGGCTGTG
BM1818	AGCTGGGAATATAACCAAAGG	AGTGCTTTCAAGGTCCATGC

Sus/Babyrousa Microsatellite

Locus	Forward primer	Reverse Primer
S0386	TCCTGGGTCTTATTTTCTA	TTTTTATCTCCAACAGTAT
S0155	TGTTCTCTGTTTCTCCTCTGTTTG	AAAGTGGAAAGAGTCAATGGCTAT
SW911	CTCAGTTCTTTGGGACTGAACC	CATCTGTGGAAAAAAAAAGCC
S0215	TAGGCTCAGACCCTGCTGCAT	TGGGAGGCTGAAGGATTGGGT
S0214	CCCTGCAAGCGTTCATCTCA	CCCTGCAAGCGTTCATCTCA
S0026	AACCTTCCCTTCCCAATCAC	CACAGACTGCTTTTTACTCC
S0149	ATTGGCTCATGAACCACCATC	GAGTTACTAATTGCCTCAGAG
S0228	GGCATAGGCTGGCAGCAACA	AGCCCACCTCATCTTATCTACACT
SW72	ATCAGAACAGTGCGCCGT	TTTGAAAATGGGGTGTTTCC
SW632	TGGGTTGAAAGATTTCCCAA	GGAGTCAGTACTTTGGCTTGA
SW951	TTTCACAACTCTGGCACCAG	GATCGTGCCCAAATGGAC
SW857	TGAGAGGTCAGTTACAGAAGACC	GATCCTCCTCCAAATCCCAT
SW936	TCTGGAGCTAGCATAAGTGCC	GTGCAAGTACACATGCAGGG
SW240	AGAAATTAGTGCCTCAAATTGG	AAACCATTAAGTCCCTAGCAAA

Anoa mtDNA

Locus	**Name**	**Sequence**	**F/R**	**Reference**
d-loop	AN4	GGTAATGTACATAACATTAATG	F	Cymbron 1999
d-loop	AN3	CGAGATGTCTTATTTAAGAGG	R	Cymbron 1999
d-loop	BethBigF-ww	ACMCCCAAAGCTGAAGTTCT	F	This study
d-loop	A-DL-R2c	GGTTGCTGGTTTCACGCGG	R	This study
Cyt-B	mta	CTCCCAGCCCCATCCAACATCTCAGCATGATGAAACTTCG	F	Schreiber 1999
Cyt-B	mtb	TTGTGATTACTGTAGCACCTCAAAATGATATTTGTCCCTCA	R	Schreiber 1999
Cyt-B	A-CB-F2a	GCCACAGCATTTATAGGATACG	F	This study
Cyt-B	A-CB-R2a	GATCGTARGATTGCGTATGC	R	This study

Sus/Babyrousa mtDNA

S. celebensis				
Locus	**Name**	Sequence	**F/R**	**Reference**
d-loop	L15387	CTCCGCCATCAGCACCCAAAG	F	Larson 2005
d-loop	H764	TGCTGGTTTCACGCGGCA	R	Larson 2005
d-loop	L119n	ATTATTRATCGTACATAGCAC	F	Larson 2007
d-loop	H16108n	GCACCTTGTTTGGATTRTCG	R	Larson 2007

Babirussa				
d-loop	L15387	CTCCGCCATCAGCACCCAAAG	F	Larson 2005
d-loop	H648n	GCTYATATGCATGGGGACT	R	Larson 2007
d-loop	BabyF	TGTACGCCAAAACATCAAGTAC	F	This study
d-loop	RuminR	GGGCGATTTTAGGTGAGATGG	R	This study

Table S7: Marginal likelihood of molecular clock analyses under different models

Clock model	Tree prior	Marginal likelihood for bovid data set	Marginal likelihood for suid data set
Strict	Constant size coalescent	-3283.86	-5861.07
Strict	Skyline coalescent	**-3261.51**	**-5847.15**
Strict	Birth-death process	-3277.08	-5857.65
Relaxed	Constant size coalescent	-3281.97	-5856.66
Relaxed	Skyline coalescent	-3261.53	-5851.94
Relaxed	Birth-death process	-3280.03	-5863.33

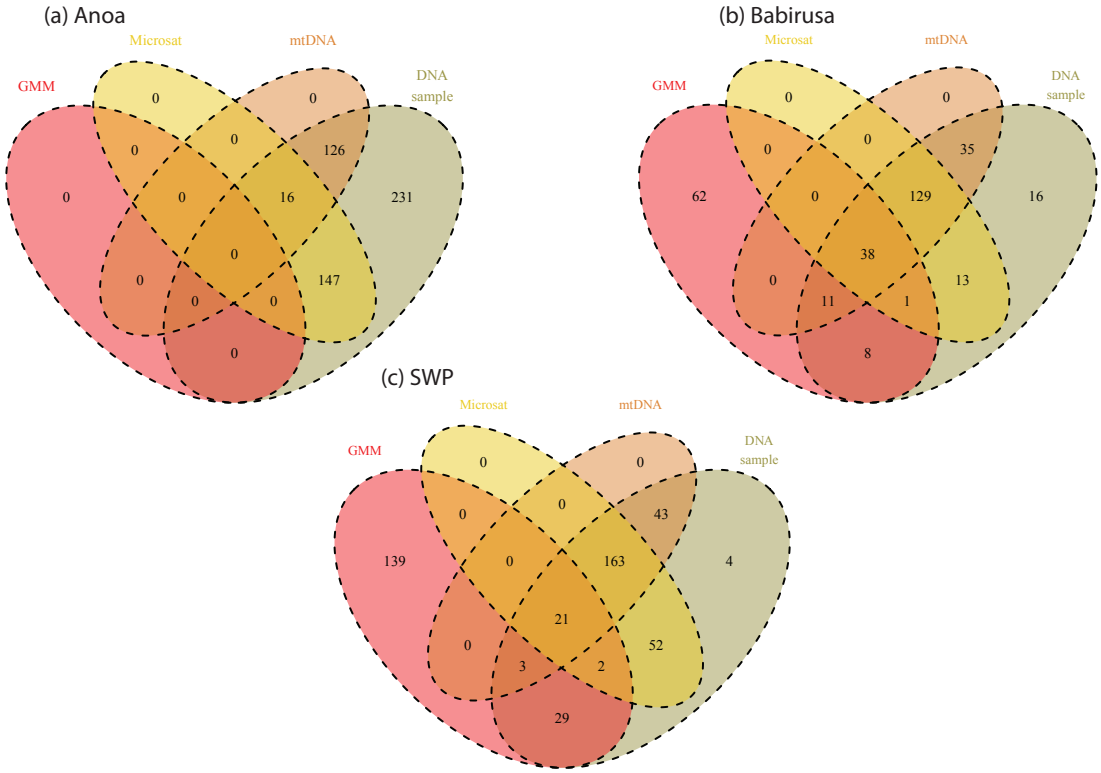


Figure S1. Venn diagram representing the number of individuals and the overlap between the various databases generated for this project. a. Anoa **b.** Babirusa **c.** *Sus celebensis.*

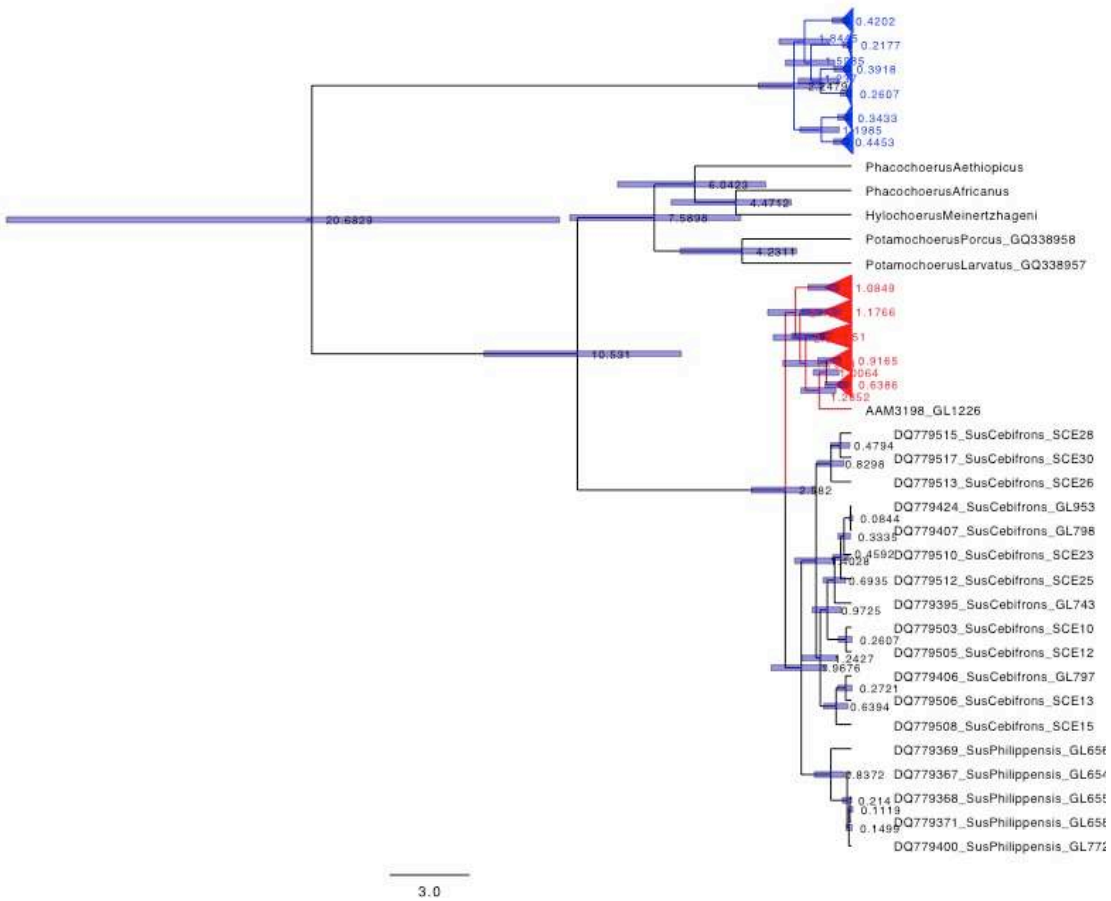


Figure S2: Molecular clock results for suids alignment

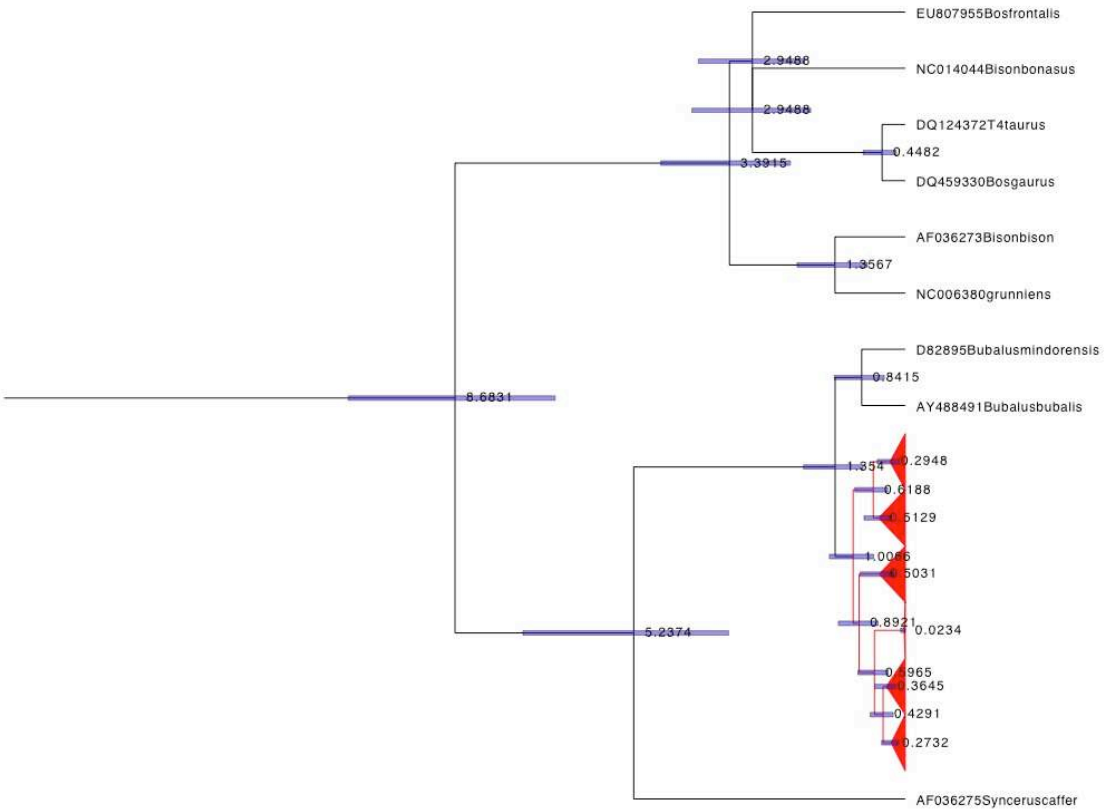


Figure S3: Molecular clock results for bovids alignment

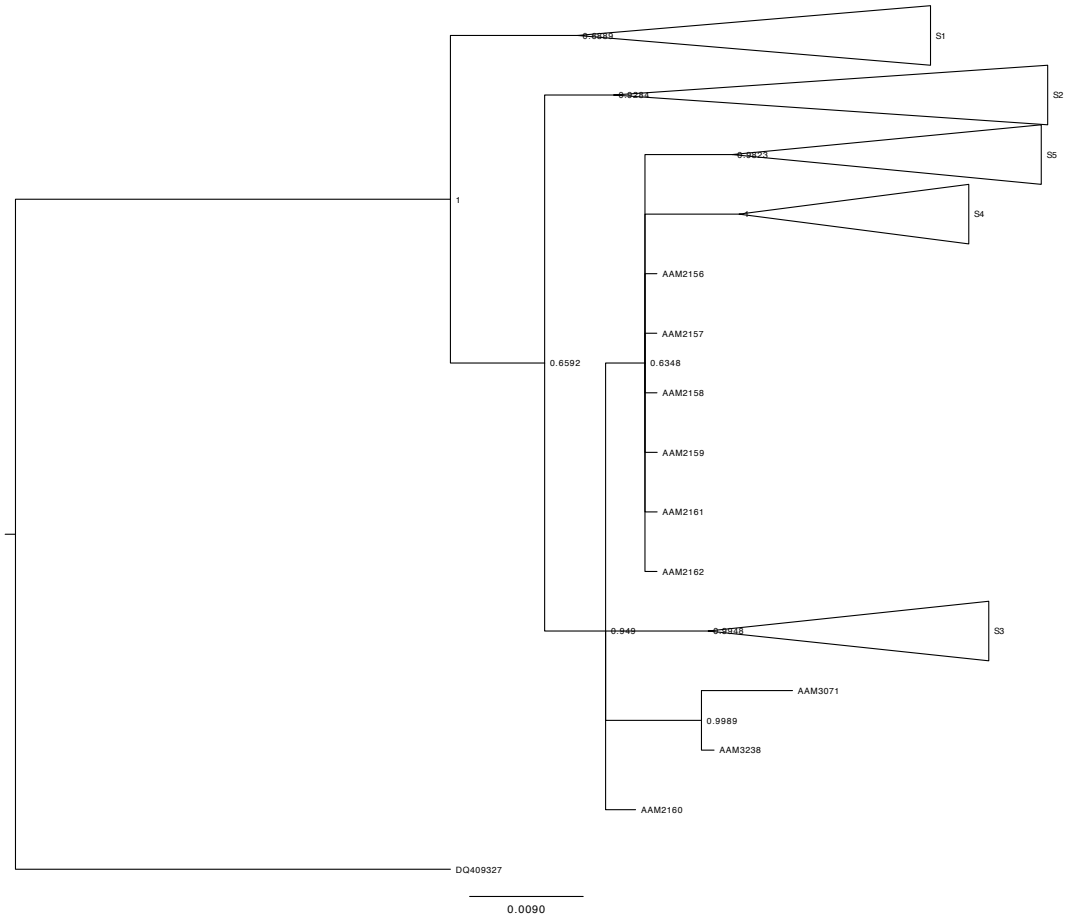


Figure S4: Bayesian phylogeny inferred from mtDNA from *Sus celebensis*. Support values represent posterior probabilities, S1-5 label represent haplogroups plotted in Fig. 1.

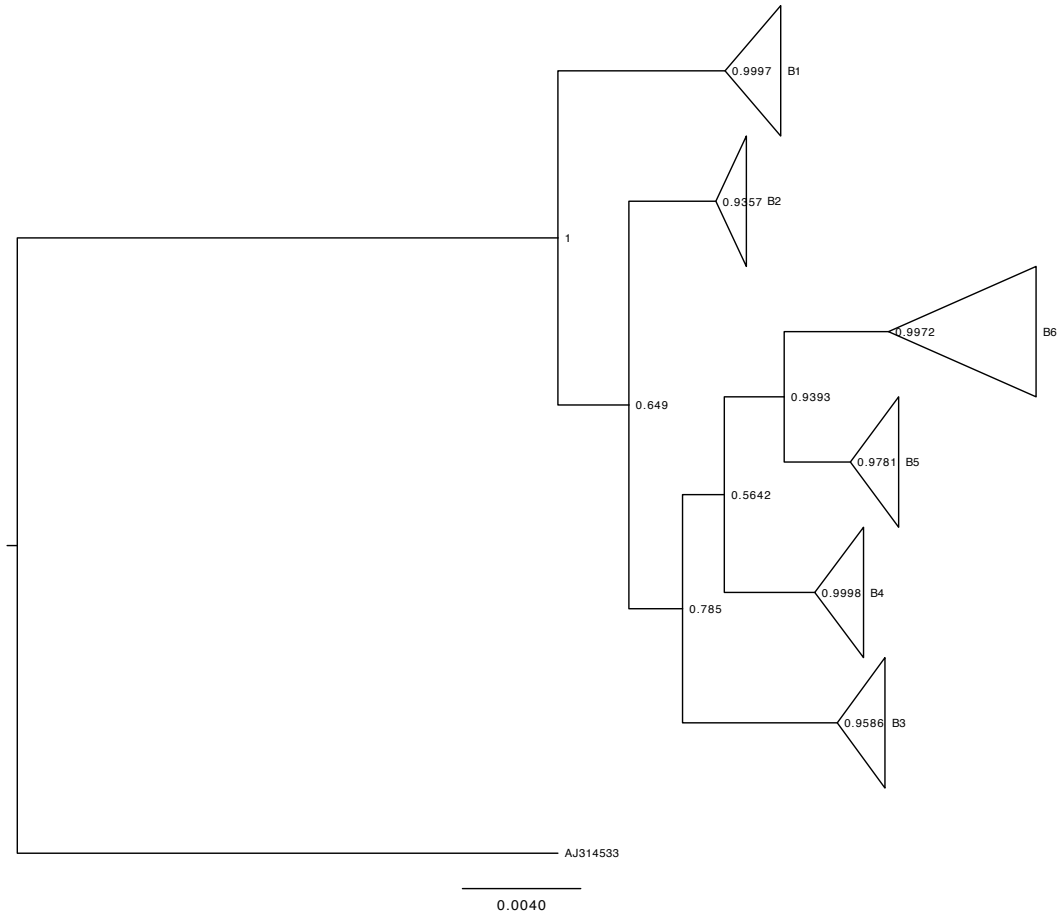


Figure S5: Bayesian phylogeny based on mtDNA from Babirusa. Support values represent posterior probabilities; B1-6 labels represent haplogroups plotted in Fig. 1.

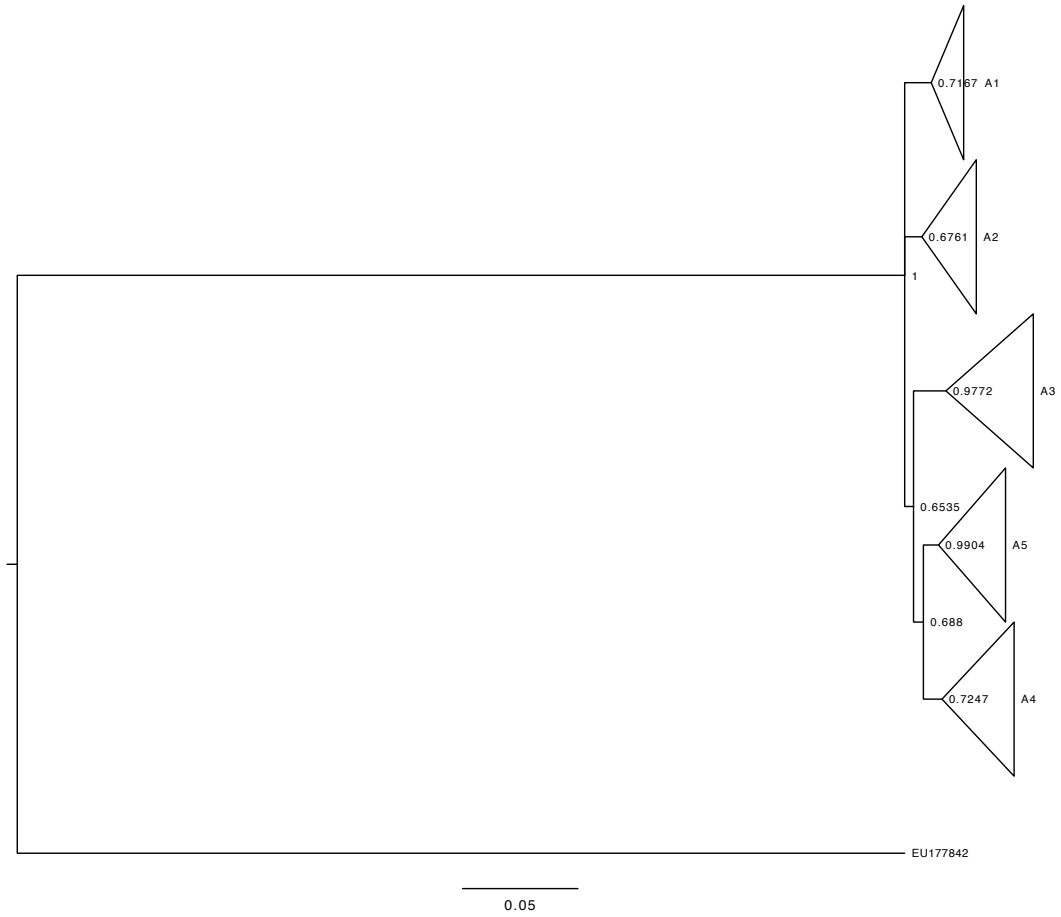
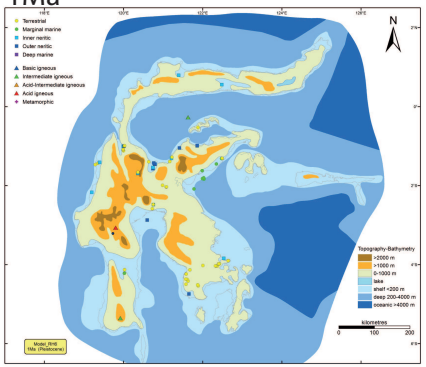
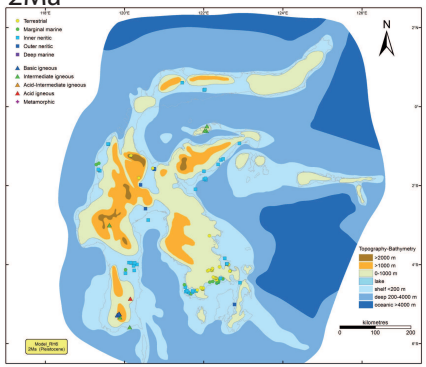


Figure S6: Bayesian phylogeny based on mtDNA from Anoa. Support values represent posterior probabilities; A1-5 labels represents haplogroups plotted in Fig. 1.

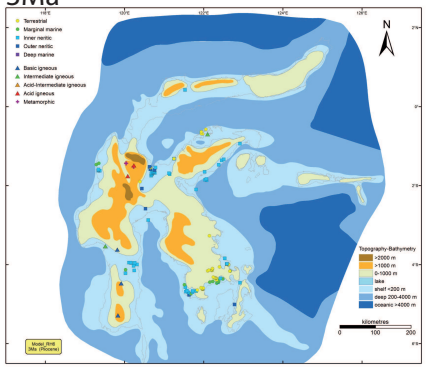
1Ma



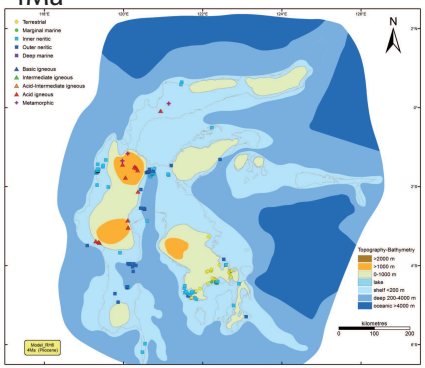
2Ma



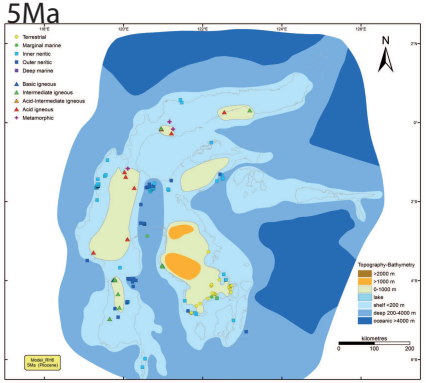
3Ma



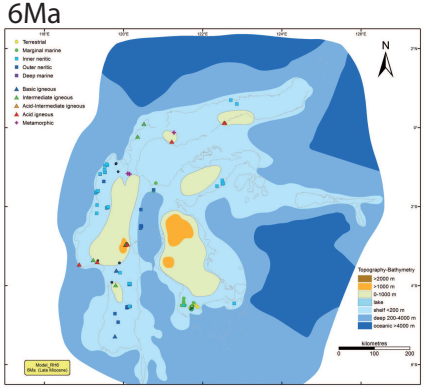
4Ma



5Ma



6Ma



8Ma

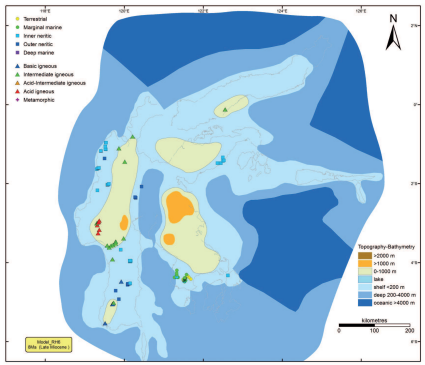
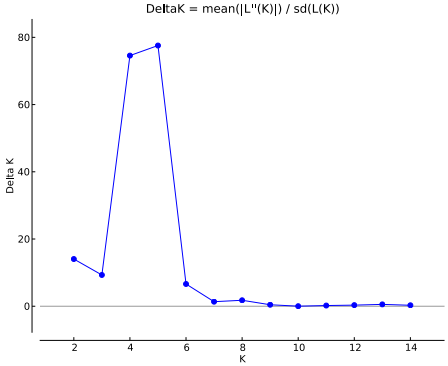
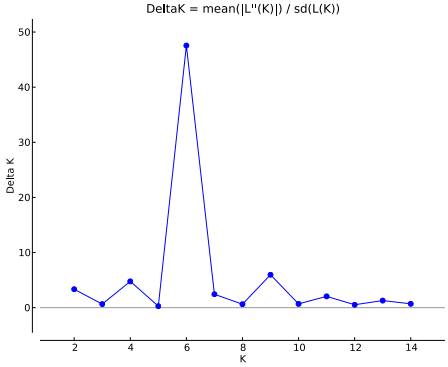


Figure S7: Tectonic reconstruction of Sulawesi over the last 8My in 1My increments.

a.



b.



c.

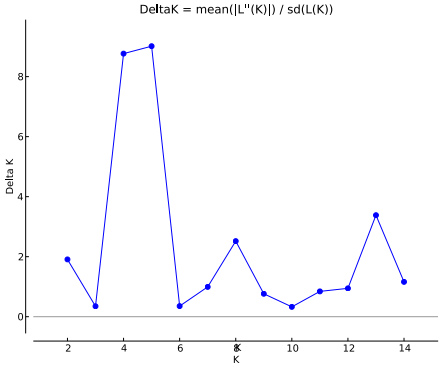


Figure S8: Δ*K* values for each species (best number of clusters in the microsatellite data). a. Anoa **b.** Babirusa **c.** Sulawesi warty pig.

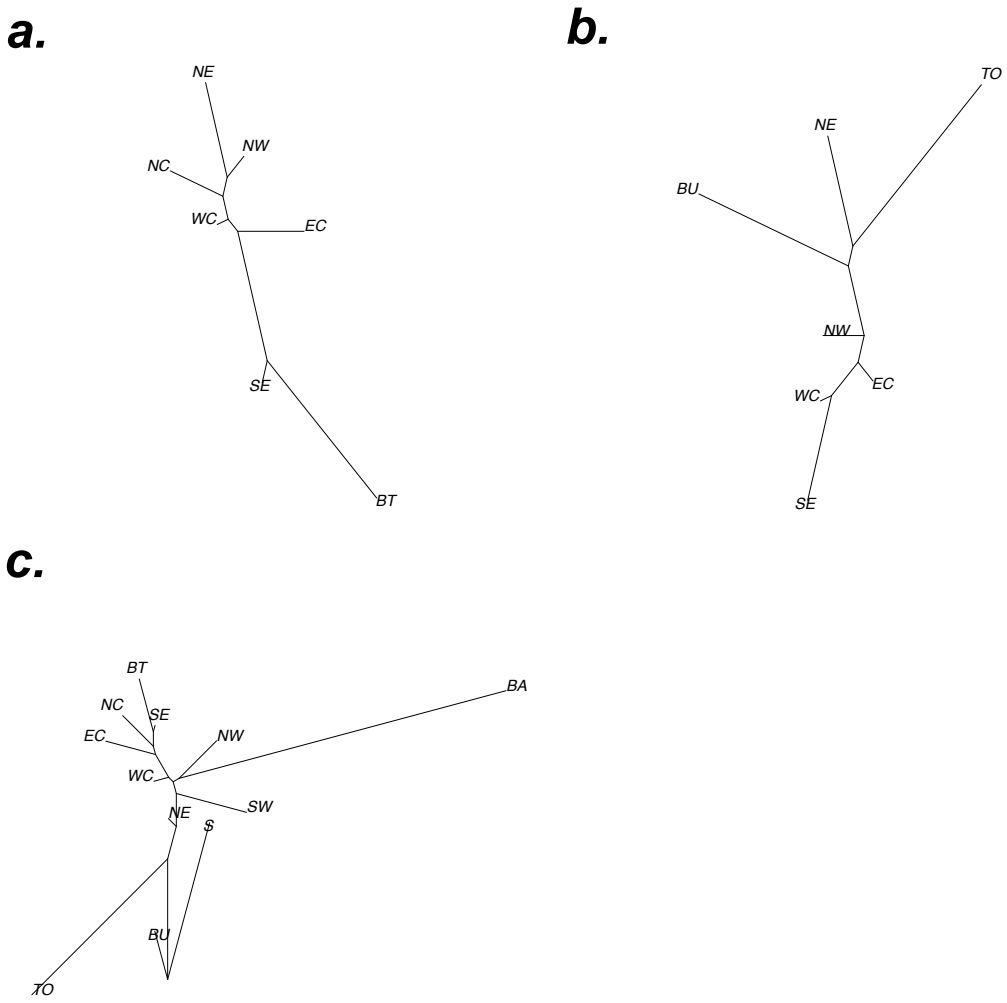


Figure S9: Neighbour-joining trees based on Fst. a. Anoa **b.** Babirusa **c.** Sulawesi warty pig.

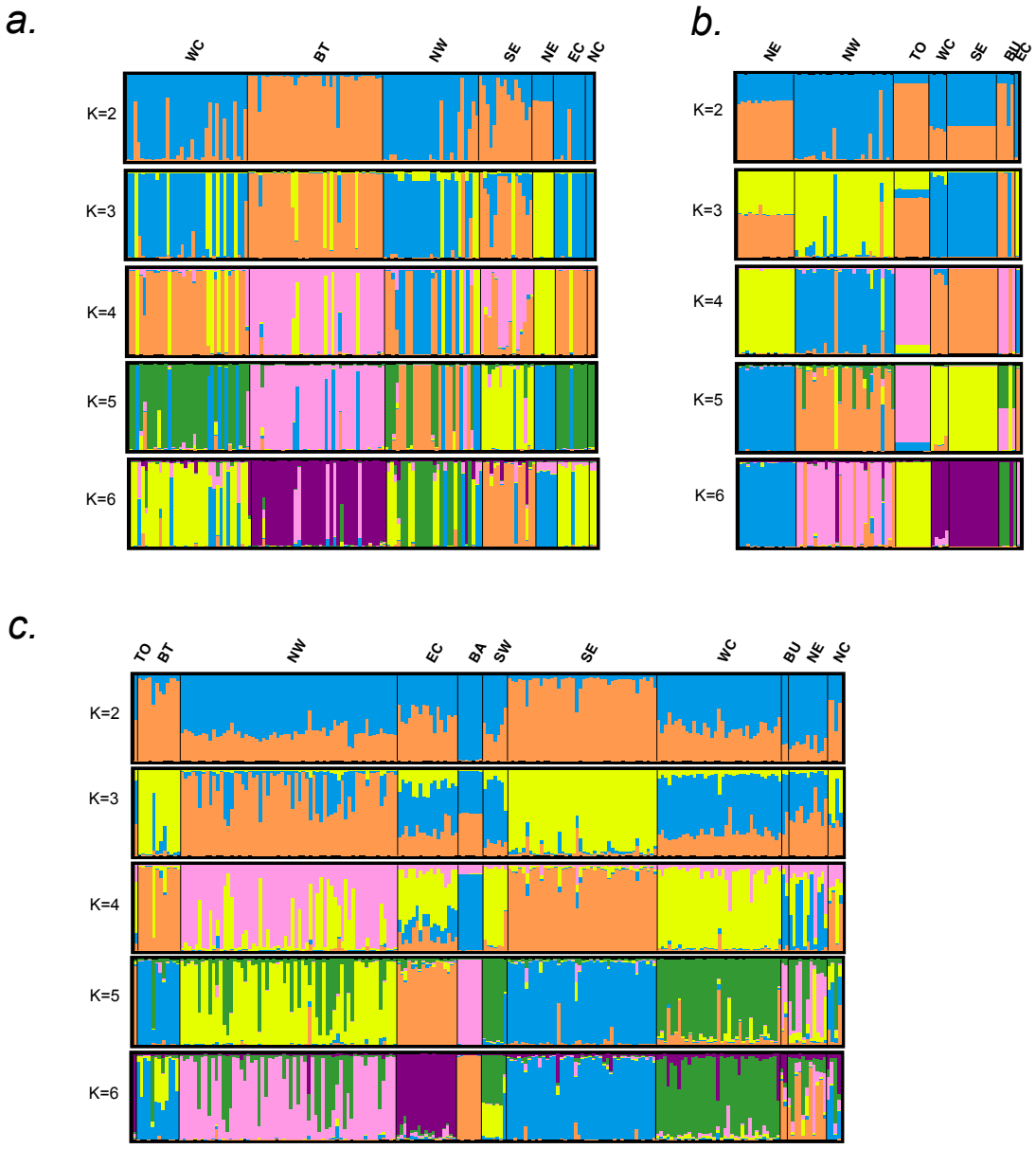


Figure S10: Results of the STRUCTURE analysis for *K*=2 to *K*=6. a. Anoa **b.** Babirusa **c.** Sulawesi warty pig.

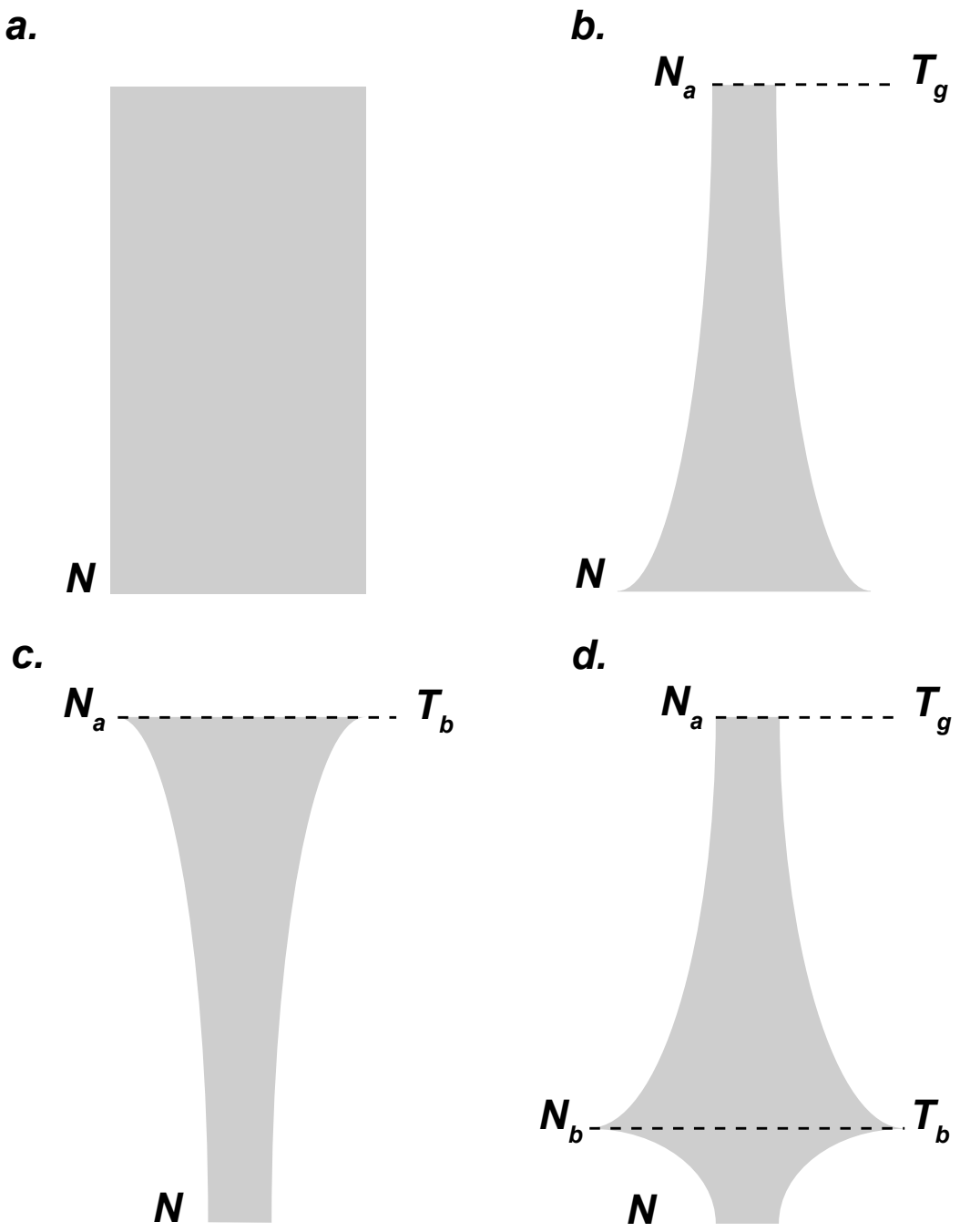


Figure S11: Various models tested using approximate Bayesian computation. **a.** Constant population size (Model 1). **b.** Population expansion (Model 2). **c.** Population bottleneck (Model 3). **d.** Population expansion followed by a bottleneck (Model 4).

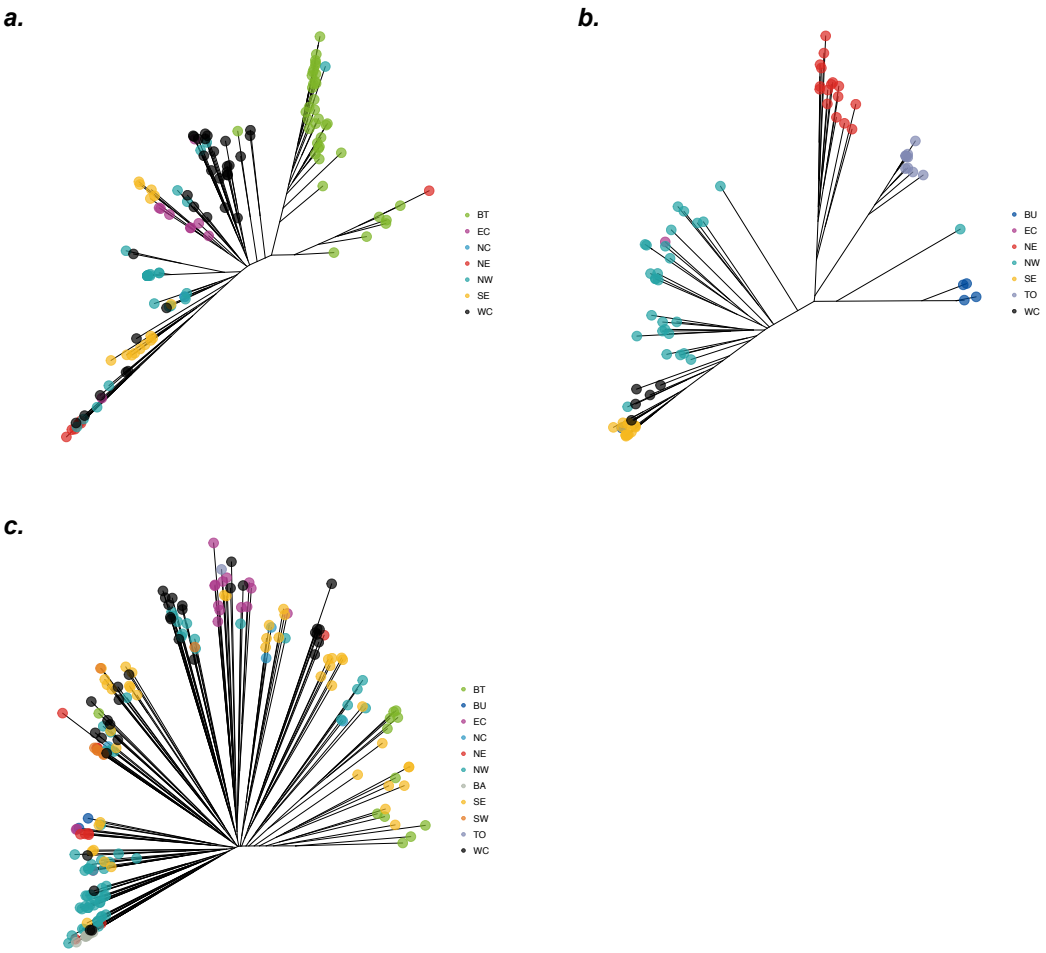


Figure S12: Neighbour-joining tree based on pairwise proportion of shared alleles using the microsatellite data. a. Anoa **b.** Babirusa **c.** Sulawesi warty pig.

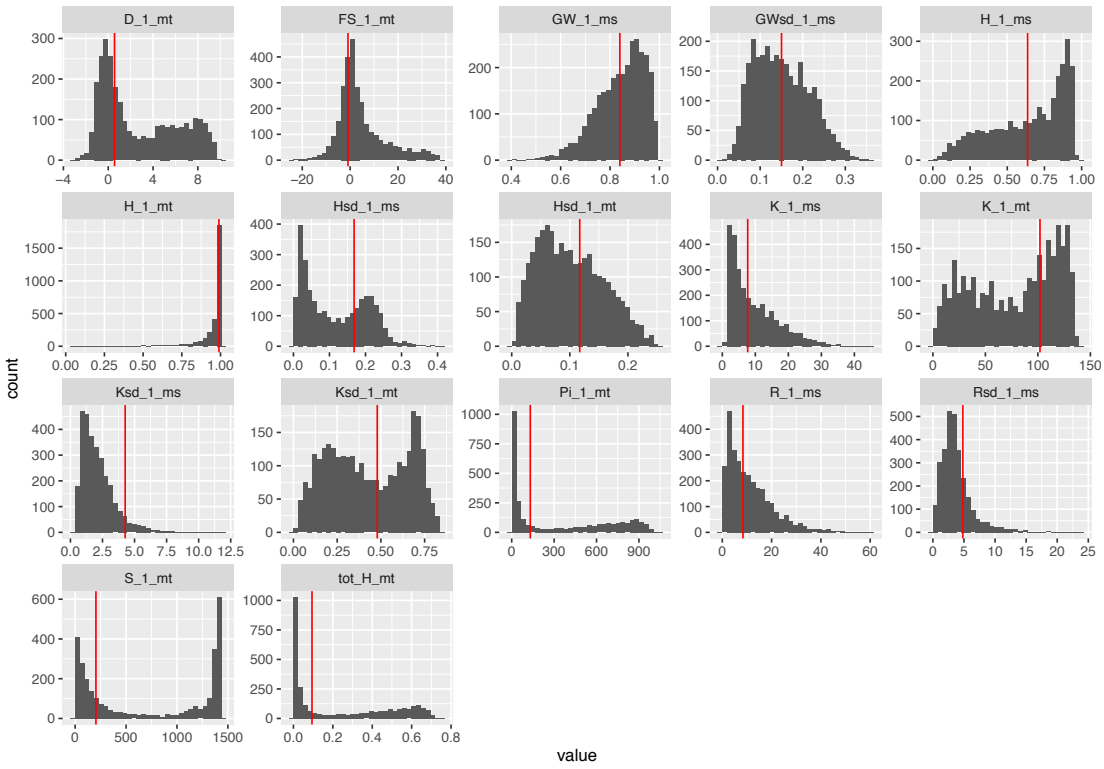


Figure S13 Observed (red vertical line) and simulated (histogram) of all summary statistics used in the approximate Bayesian computation analysis (Anoa).

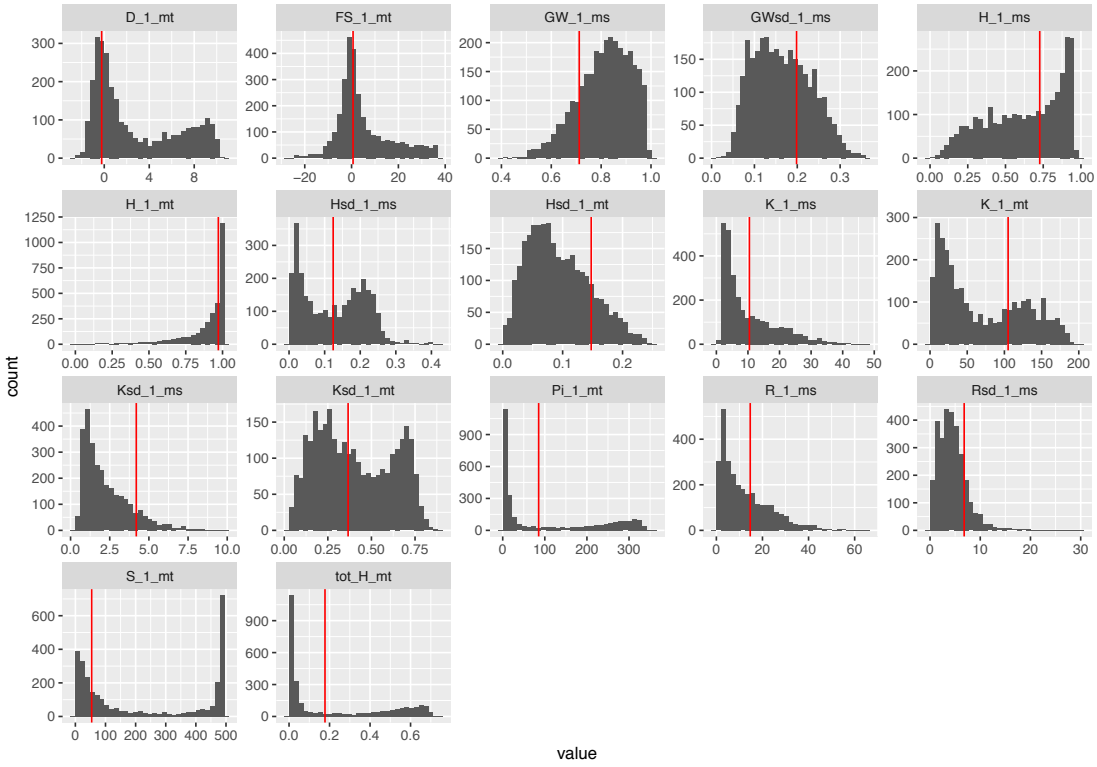


Figure S14 Observed (red vertical line) and simulated (histogram) of all summary statistics used in the approximate Bayesian computation analysis (Babirusa).

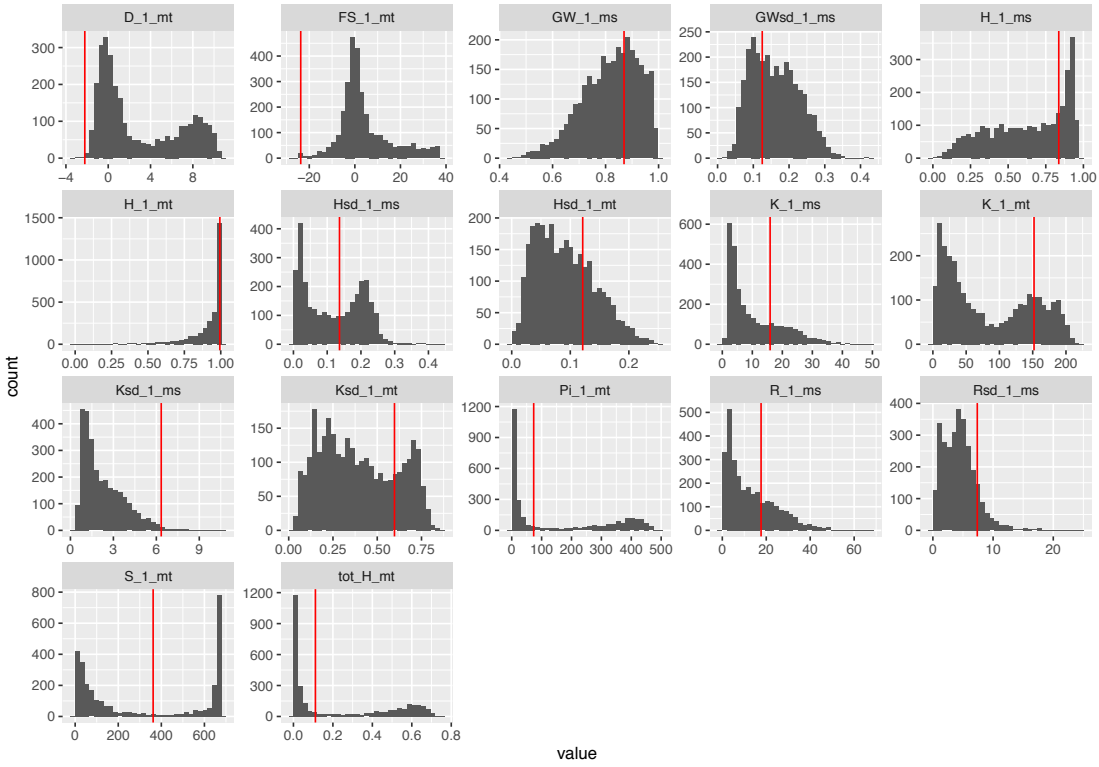


Figure S15 Observed (red vertical line) and simulated (histogram) of all summary statistics used in the approximate Bayesian computation analysis (SWP).

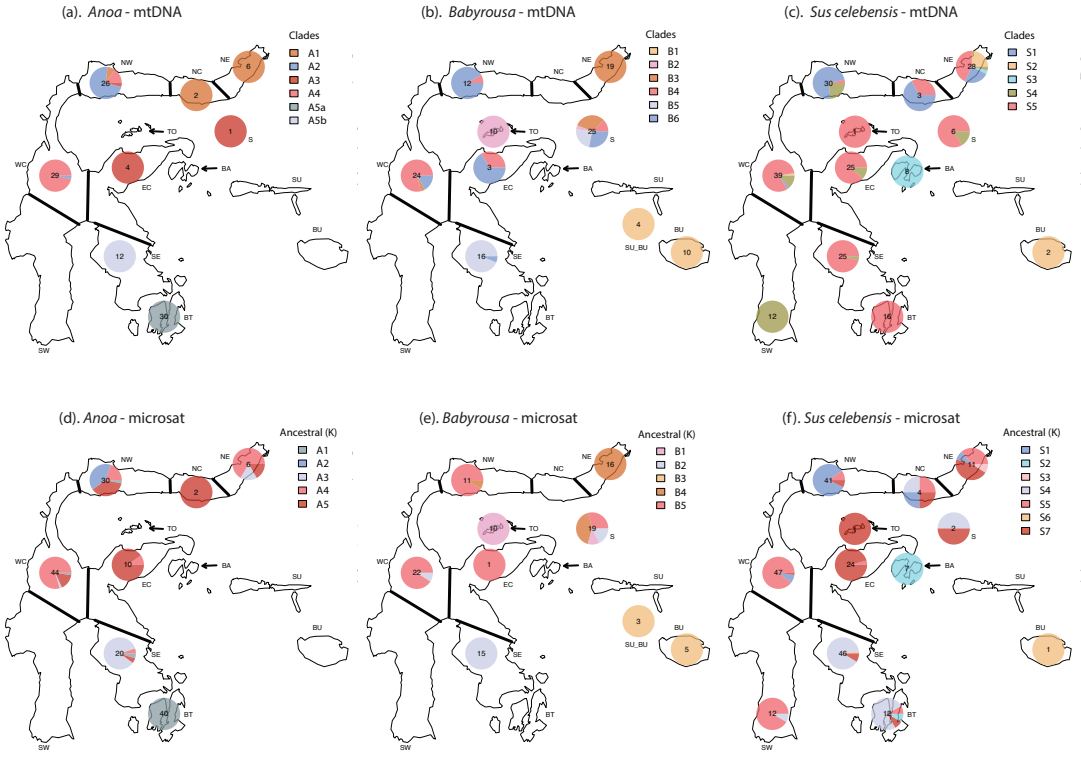


Figure S16: Population structure of each species inferred from mtDNA, microsatellites. a. to c., Proportion of haplogroups in each region of endemism and phylogeny of Anoa (**a.**), Babirusa (**b.**) and Sulawesi warty pig (**c.**). Numbers in pie charts represent the sample size in a given region. **d. to f.,** Result of the STRUCture analysis using the microsatellite data plotted on the map and as a bar chart (Fig. S10) for Anoa (**d.**), Babirusa (**e.**) and SWP (**f.**). The best *K* value for each species was used (*K*=5 for Anoa; *K*=6 for Babirusa; *K*=5 for SWP). NE=North East; NC=North Central; NW=North West; TO=Togian; BA=Banggai Archipelago; EC=East Central; WC=West Central; SU=Sula; BU=Buru; S=Sula or Buru; SE=South East; SW= South West; BT=Buton.